

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

**Contributions to improve the power,
efficiency and scope of control-chart
methods**

A thesis submitted in partial fulfilment of the
requirements for the degree of

Doctor of Philosophy

in

Statistics

at Massey University, Albany, New Zealand

Nurudeen Adedayo Adegoke

August 2019

Copyright is owned by the Author of the thesis. Permission is given for a copy to be downloaded by an individual for the purpose of research and private study only. The thesis may not be reproduced elsewhere without the permission of the Author.

Abstract

Detection of outliers and other anomalies in multivariate datasets is a particularly difficult problem which spans across a range of systems, such as quality control in factories, microarrays or proteomic analyses, identification of features in image analysis, identifying unauthorized access in network traffic patterns, and detection of changes in ecosystems. Multivariate control charts (MCC) are popular and sophisticated statistical process control (SPC) methods for monitoring characteristics of interest and detecting changes in a multivariate process. These methods are divided into memory-less and memory-type charts which are used to monitor large and small-to-moderate shifts in the process, respectively. For example, the multivariate χ^2 is a memory-less control chart that uses only the most current process information and disregards any previous observations; it is typically used where any shifts in the process mean are expected to be relatively large. To increase the sensitivity of the multivariate process control tool for the detection of small-to-moderate shifts in the process mean vector, different multivariate memory-type tools that use information from both the current and previous process observations have been proposed. These tools have proven very useful for multivariate independent normal or “nearly” normal distributed processes.

Like most univariate control-chart methods, when the process parameters (i.e., the process mean vector or covariance parameters, or both) are unknown, then MCC methods are based on

estimated parameters, and their implementation occurs in two phases. In Phase I (retrospective phase), a historical reference sample is studied to establish the characteristics of the in-control state and evaluate the stability of the process. Once the in-control reference sample has been deemed to be stable, the process parameters are estimated from Phase I, and control chart limits are obtained for use in Phase II. The Phase II aspect initiates ongoing regular monitoring of the process. If successive observed values obtained at the beginning of Phase II fall within specified desired in-control limits, the process is considered to be in control. In contrast, any observed values during Phase II which fall outside the specified control limits indicate that the process may be out of control, and remedial responses are then required.

Although conventional MCC are well developed from a statistical point of view, they can be difficult to apply in modern, data-rich contexts. This serious drawback comes from the fact that classical MCC plotting statistics requires the inversion of the covariance matrix, which is typically assumed to be known. In practice, the covariance matrix is seldom known and often empirically estimated, using a sample covariance matrix from historical data. While the empirical estimate of the covariance matrix may be an unbiased and consistent estimator for a low-dimensional data matrix with an adequate prior sample size, it performs inconsistently in high-dimensional settings. In particular, the empirical estimate of the covariance matrix can lead to inflated false-alarm rates and decreased sensitivity of the chart to detect changes in the process.

Also, the statistical properties of traditional MCC tools are accurate only if the assumption of multivariate normality is satisfied. However, in many cases, the underlying system is not multivariate normal, and as a result, the traditional charts can be adversely affected. The necessity of this assumption generally restricts the application of traditional control charts to monitoring

industrial processes.

Most MCC applications also typically focus on monitoring either the process mean vector or the process variability, and they require that the process mean vector be stable, and that the process variability be independent of the process mean. However, in many real-life processes, the process variability is dependent on the mean, and the mean is not necessarily constant. In such cases, it is more appropriate to monitor the coefficient of variation (CV). The univariate CV is the ratio of the standard deviation to the mean of a random variable. As a relative dispersion measure to the mean, it is useful for comparing the variability of populations having very different process means. More recently, MCC methods have been adapted for monitoring the multivariate coefficient of variation (CV). However, to date, studies of multivariate CV control charts have focused on power—the detection of out-of-control parameters in Phase II, while no study has investigated their in-control performance in Phase I. The Phase I data set can contain unusual observations, which are problematic as they can influence the parameter estimates, resulting in Phase II control charts with reduced power. Relevant Phase I analysis will guide practitioners with the choice of appropriate multivariate CV estimation procedures when the Phase I data contain contaminated samples.

In this thesis, we investigated the performance of the most widely adopted memory-type MCC methods: the multivariate cumulative sum (MCUSUM) and the multivariate exponentially weighted moving average (MEWMA) charts, for monitoring shifts in a process mean vector when the process parameters are unknown and estimated from Phase I (chapters 2 and 3). We demonstrate that using a shrinkage estimate of the covariance matrix improves the run-length performance of these methods, particularly when only a small Phase I sample size is available. In chapter 4,

we investigate the Phase I performance of a variety of multivariate CV charts, considering both diffuse symmetric and localized CV disturbance scenarios, and using probability to signal (PTS) as a performance measure.

We present a new memory-type control chart for monitoring the mean vector of a multivariate normally distributed process, namely, the multivariate homogeneously weighted moving average (MHWMA) control chart (chapter 5). We present the design procedure and compare the run length performance of the proposed MHWMA chart for the detection of small shifts in the process mean vector with a variety of other existing MCC methods. We also present a dissimilarity-based distribution-free control chart for monitoring changes in the centroid of a multivariate ecological community (chapter 6). The proposed chart may be used, for example, to discover when an impact may have occurred in a monitored ecosystem, and is based on a change-point method that does not require prior knowledge of the ecosystem's behaviour before the monitoring begins. A novel permutation procedure is employed to obtain the control-chart limits of the proposed charting test-statistic to obtain a suitable distance-based model of the target ecological community through time.

Finally, we propose enhancements to some classical univariate control chart tools for monitoring small shifts in the process mean, for those scenarios where the process variable is observed along with a correlated auxiliary variable (chapters 7 through 9). We provide the design structure of the charts and examine their performance in terms of their run length properties. We compare the run length performance of the proposed charts with several existing charts for detecting a small shift in the process mean. We offer suggestions on the applications of the proposed charts (in chapters 7 and 8), for cases where the exact measurement of the process variable of interest or the auxiliary

variable is difficult or expensive to obtain, but where the rank ordering of its units can be obtained at a negligible cost.

Thus, this thesis, in general, will aid practitioners in applying a wider variety of enhanced and novel control chart tools for more powerful and efficient monitoring of multivariate process. In particular, we develop and test alternative methods for estimating covariance matrices of some useful control-charts' tools (chapters 2 and 3), give recommendations on the choice of an appropriate multivariate CV chart in Phase I (chapter 4), present an efficient method for monitoring small shifts in the process mean vector (chapter 5), expand MCC analyses to cope with non-normally distributed datasets (chapter 6) and contribute to methods that allow efficient use of an auxiliary variable that is observed and correlated with the process variable of interest (chapters 7 through 9).

Acknowledgments

Alhamdulillah, all praises and gratitude are due to The Almighty Allah for His countless and endless blessings, not the least of which has been enabling me to complete this thesis. May His peace and blessings be upon His beloved Messenger.

This thesis could not have been completed without the financial support provided by the School of Natural and Computational Sciences (SNCS), Massey University, Auckland, through a graduate teaching assistantship position awarded to me, for which I am very grateful. Also, I am thankful to the School for providing me with so many resources that enhance the smooth running of my research.

I am particularly grateful to my primary supervisor Dr. Matthew D.M. Pawley, for his continuous intellectual guidance, support, positive feedback and advice throughout the research journey. I am very grateful to him for having trust in my abilities, giving me the maximum freedom to attempt whatever I desired in the research, and for being accessible to discuss any research related issues. I feel very fortunate to have Distinguished Professor Marti J. Anderson as one of my co-supervisors, her support, incisive comments, constructive discussions, and suggestions will continue to guide me throughout my future career. I also thank her for offering me my first professional job in New Zealand. I am also very grateful to my co-supervisor Dr. Adam N.H. Smith

for his support, insightful and detailed responses throughout the research journey. I gratefully acknowledge the valuable feedback received from him to improve the thesis. Also, for making his office available for our weekly meetings. Without the inputs and help from all of my supervisors, this research would not have been possible. I look forward to working with you all in the future.

I offer recognition to Associate Prof. Saddam Akber Abbassi (Department of Mathematics, Statistics and Physics, Qatar University, Doha, Qatar), and Prof. Muhammad Riaz (Mathematics and Statistics Department, King Fahd University of Petroleum and Minerals, KSA), for their help and support during my research. They are great people and I always learned a lot from them through fruitful discussions. I am sincerely grateful to them and I look forward to maintaining working with them in the future.

I would like to express my sincere gratitude to a number of currents and past SNCS staff and students including the wonderful administrative staff (Annette Warbrooke, Anil Malhotra, Freda Mickisch, Lyn Shave, Linh Mills, and Jukai Zhou), the following academic staff (Dr. Beatrix Jones, Dr. Barry McDonald, Dr. David Aguirre, and Dr. Libby Liggins), and various office mates (Dr. David Acuna Marrero, Amir Bashir, Sheng Gong, Christian Bläsche, Sidra Zafar, and Cong Yao), who have encouraged, helped and supported me in any form during my research.

Throughout the academics, professional careers and life, I have been fortunate to have the assistance, guidance, and support of many individuals including my mentors and teachers (Dr. Dauda Taofik Oyedele, Associate Prof. Femi B. Adebola, Associate Prof. Fasoranbaku A. Olu-soga, Akintoye Nurain Akinjide, Akinsola Egbelakin, Associate Prof. Temitope Egbelakin, Dr. AbdulRasheed Amidu, Taoheed Raji, Alhaji Ismail Jayeiola, and Adegbite Fatimah), my friends and colleagues (Azeez Omitogun, Idowu Afis Oladele, Jimoh Amodu, Fatoyinbo Hamed, Ajayi

Adekunle, Olasunkanmi Kehinde, Rufai Muftau, Adelabu Isaiah, Adeleke Kabir, Rhaji Idowu, Afeez Abidemi, Ayooluwa Odufowora, Sanusi Ridwan, Akintoye Oluwasegun, and Muhammed Bolomope), my siblings (Amudalat, Basirat, Usman, Abdul-Sallam, and Nuriat), and my lovely parents (my daddy (Jimoh A. Adegoke), and my late mum (Rasheedat Adegoke)), my step-mum (Muriliat Adegoke)), my parents-in-law, my families, and my other teachers, who have helped me to stand my ground and have the kind of career that I can be proud of. You have all inspired me to pursue my goals with hard work and dedication. It would be impossible to count all the ways that you have helped me in my career. Thank you so much for all that you have done and your prayers; I only hope I can reciprocate the support sometime in the future.

To my beloved children (Nadrah and Naeem), I am so proud of you and honoured to be your father. Your arrival during my Ph.D gave me more inner strengths to work harder and be more focused. Importantly, I want to express my deepest gratitude to my lovely wife, Khadijat Ganiyu, for your consistent support, patience and countless sacrifices that help me get to this point. Thank you for taking care of our son (Naeem), daughter (Nadrah) and myself, and for cheering me up when I feel tired. You are wonderful and I could not have finished my Ph.D without the support you provided. But most of all, thank you for being my best friend. I owe you everything, and I wholeheartedly dedicate my Ph.D to you and The Almighty Allah.

Contents

Abstract	i
Acknowledgments	vi
List of Tables	xvi
List of Figures	xxiii
1 Introduction	1
1.1 Thesis Contribution	6
1.2 Thesis Outline	7
2 Multivariate cumulative sum control charts for individual-observations monitoring when parameters are estimated	11
2.1 Introduction	12
2.2 Multivariate cumulative sum control charts	15
2.3 Estimates of the mean vector and covariance matrix	18
2.4 Shrinkage estimate of the covariance matrix	19

2.5	Assessing the in-control performance of MCUSUM and MCI control charts using simulation	21
2.5.1	Simulation method	21
2.5.2	Simulation results of the in-control ARL	23
2.6	Corrected limits of the MCUSUM and MCI charts	25
2.7	Out-of-control ARL performance using corrected limits	28
2.8	Estimation of the corrected limits from the shrinkage estimate	32
2.9	Illustrative example	33
2.10	General conclusions	36
3	Multivariate exponentially weighted moving average (MEWMA) control charts for individual-observations monitoring when parameters are estimated	38
3.1	Introduction	39
3.2	Multivariate exponentially weighted moving average control chart (MEWMA) . . .	42
3.2.1	General framework	42
3.3	Estimating the Phase I parameters	44
3.3.1	Estimating the mean vector	45
3.3.2	Estimating the covariance matrix	45
3.4	Simulation study of the performance of MEWMA control charts	51
3.4.1	Methods	51
3.4.2	Results of the simulation study	53
3.5	Illustrative examples	63
3.5.1	Example 1: Bimetal thermostat dataset	64

3.5.2	Example 2: Gene expression dataset	64
3.6	General conclusions	69
4	Multivariate coefficient of variation (CV) control charts in Phase I of statistical process control (SPC)	71
4.1	Introduction	72
4.2	Multivariate CV in Phase I	76
4.3	Performance evaluation in Phase I	78
4.4	Results and discussion	80
4.5	Illustrative example	83
4.6	Summary and conclusion	91
5	A new multivariate control chart based on homogeneously weighted moving average	93
5.1	Introduction	94
5.2	Literature review: the memory-type control charts	96
5.2.1	The MCUSUM chart	96
5.2.2	The MCI chart	97
5.2.3	The MEWMA chart	98
5.3	The multivariate homogeneously weighted moving average (MHWMA) control chart	100
5.4	Performance evaluation	102
5.5	Average run length comparisons	106
5.6	Illustrative example	106
5.6.1	Simulated example	107

5.6.2	Bimetal thermostat dataset example	110
5.7	Conclusion	114
6	A new distribution-free multivariate control chart for ecological applications	115
6.1	Introduction	116
6.1.1	Background methodologies	120
6.1.2	The multivariate chi-square chart	120
6.2	Description of the proposed method	122
6.2.1	Parametric case	122
6.2.2	Non-parametric case	124
6.2.3	Progressive analysis	128
6.3	Distance to centroid approach	129
6.4	Performance of the new method and results	130
6.4.1	Multivariate normal simulation description and results	131
6.4.2	Ecological simulated data set and results	133
6.5	Ecological example	134
6.6	Conclusion and discussion	136
7	EWMA control chart for monitoring the mean of a process that is negatively correlated with an auxiliary variable under some ranked sampling schemes	142
7.1	Introduction	143
7.2	The ranked set sampling (RSS) and median ranked set sampling (MRSS) procedures	146
7.3	The product estimator	148
7.3.1	Efficiency of the product estimator based on the different sampling schemes .	149

7.4	The classical EWMA control chart	154
7.5	The proposed charts	156
7.5.1	The proposed $EWMA_P$ control chart based on SRS	156
7.5.2	The proposed $EWMA_{PRY}$ and $EWMA_{PRX}$ control charts based on RSS . . .	157
7.5.3	The proposed $EWMA_{PMY}$ and $EWMA_{PMX}$ control charts based on MRSS .	159
7.6	Performance measures and comparison	162
7.7	Illustrative examples	168
7.7.1	Simulated dataset	170
7.7.2	Real dataset	173
7.8	General conclusions	175
8	EWMA control chart for monitoring the mean of a process that is correlated with an auxiliary variable under some ranked sampling scheme	180
8.1	Introduction	181
8.2	The ranked set sampling and median ranked set sampling schemes	183
8.3	The proposed charts	185
8.4	Simulation study	188
8.5	Results and discussion	191
8.6	Illustrative examples	198
8.7	Summary and conclusion	199
9	Efficient monitoring of a process mean using an auxiliary variable under simple random sampling	201
9.1	Introduction	202

9.2	The AHWMA Control Chart	204
9.3	Performance assessments and comparisons	207
9.4	Robustness to non-normality of the chart	213
9.5	Step by step algorithm for constructing the AHWMA chart when parameters are unknown	216
9.6	Industrial application	220
9.7	Conclusion and Discussion	221
10	Summary and Recommendations for Future Research	224
10.1	Summary	224
10.2	Future Work	228
	Bibliography	230
	Appendices	256
A	Other research involvements:	257
A.1	Co-author papers	257
A.2	Conferences	257
B	Formulas and Derivations	259
B.1	Derivation of the mean vector and covariance matrix of H_i	259
B.2	Proof of the non-centrality parameter	260
C	R Scripts	262
C.1	R code for the corrected limits in Chapter 3	262

C.2	R code of the results (including ARL and SDRL) of the MHWMA chart in Chapter 5	266
C.3	R-codes for the analysis in Chapter 9	269
C.3.1	R-codes Classical CUSUM, EWMA and HWMA charts	269

List of Tables

1.1	The study done including the control chart methods examined, the novelty of each chapter and the size of the subgroup sample (n) considered. Some of the control chart methods examined include multivariate cumulative sum (MCUSUM), exponentially weighted moving average (EWMA), multivariate EWMA (MEWMA), homogeneously weighted moving average (HWMA), and multivariate HWMA (MHWMA).	7
2.1	The upper control limits: h_1 and h_2 , of the MCUSUM and MCI charts, respectively, that produce in-control ARL of 200 when the charts' statistics are obtained from known parameters, for different values of p and k	23
2.2	In-control ARL of the MCUSUM and MCI control charts for $p = 2$ or 5 , when different Phase I samples (m), of size $n = 1$, are used in estimating the unknown parameters. We used $k = 0.25$, and the upper control limits that give in-control ARL of 200 when the parameters are known.	24

2.3	In-control ARL of the MCUSUM and MCI control charts for $p = 2$ or 5 , when different Phase I samples (m), of size $n = 1$, are used in estimating the unknown parameters. We used $k = 0.5$, and the upper control limits that give in-control ARL of 200 when the parameters are known.	25
2.4	In-control ARL of the MCUSUM and MCI control charts for $p = 2$ or 5 , when different Phase I samples (m), of size $n = 1$, are used in estimating the unknown parameters. We used $k = 1$, and the upper control limits that give in-control ARL of 200 when the parameters are known.	26
2.5	In-control ARL of the MCUSUM and MCI control charts for $p = 5$, and some selected values of (m). Here, we used $k = 0.5$, and the upper control limits that give an in-control ARL of 500 when the parameters are known.	26
2.6	The corrected upper control limits: (h_{c1}) and (h_{c2}) , that produced an in-control ARL of 200 when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters. The value of k used here is $k = 0.25$	28
2.7	The corrected upper control limits: (h_{c1}) and (h_{c2}) , that produced an in-control ARL of 200 when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters. The value of k used is $k = 0.5$	29
2.8	The corrected upper control limits: (h_{c1}) and (h_{c2}) , that produced an in-control ARL of 200 when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters. The value of k used is $k = 1$	30
2.9	The zero-state ARL_1 of the MCUSUM and MCI charts from different methods of estimating the covariance matrix, for $p = 5$, different values of m and $k = 0.5$	31

2.10	The steady-state ARL_1 of the MCUSUM and MCI charts from different methods of estimating the covariance matrix, for $p = 5$, different values of m and $k = 0.5$	32
2.11	Simple linear models that may be used to estimate the corrected upper control limit h_{c1} that will give an in-control $ARL \approx 200$ for a shrinkage-MCUSUM control chart, for different values of p and reference parameter k . The coefficient of determination for each model is given in parentheses.	33
2.12	Simple linear models that may be used to estimate the corrected upper control limit h_{c2} that will give an in-control $ARL \approx 200$ for a shrinkage-MCI control chart, for different values of p and reference parameter k . The coefficient of determination for each model is given in parentheses.	33
3.1	Table of target matrices and their optimal shrinkage intensities. The terms $avg(.)$ and $Med(.)$ (in Equations (3.17) and (3.18)) are used to denote the mean and median of the sample variances across all variables, respectively.	50
3.2	The upper control limit, h , that produces $AARL_0$ of 200 when the MEWMA chart statistic is obtained from known parameters, for different values of p and r	53
3.3	$AARL_0$ values when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters under independent covariance structure. The $AARL_0$ is fixed at 200, and $r = 0.1$	57
3.4	$AARL_0$ values when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters under $AR(1)$ covariance structure. The $AARL_0$ is fixed at 200, and $r = 0.1$	58

3.5	The corrected upper control limit (h_c), that produced $AARL_0$ of 200 when different Phase I samples (m), of size $n = 1$ are used to estimate the unknown parameters.	60
3.6	The Modified Average Quality Loss (MAQL) for control charts using various methods of estimating the covariance matrix. Results are shown for dimensionality $p = 2, 3, 4$ or 5 and smoothing parameter $r = 0.05$. The corrected limits presented in Table 3.5, which standardized the $AARL_0$ to 200, were used.	63
4.1	The $d_2(n)$ and V_n values for different values of m and n for all the three charts when $\alpha = 0.01$	80
4.2	The PTS values of the localized CV disturbances model for $m = 30$, $\alpha = 0.01$ at different levels of n and m_1	82
4.3	The PTS values of the localized CV disturbances model for $m = 75$, $\alpha = 0.01$ at different levels of n and m_1	83
4.4	The PTS values of the diffused symmetric CV disturbances model for $m = 30$, $\alpha = 0.01$ at different levels of n and b	84
4.5	The PTS values of the diffused symmetric CV disturbances model for $m = 75$, $\alpha = 0.01$ at different levels of n and b	85
4.6	Sample information and CV estimators for the carbon fiber tubing data	88
4.7	Regression test results for the carbon fiber tubing data	90
5.1	ARL Values for MHWMA Charts ($p = 2$).	103
5.2	ARL values for MHWMA charts ($w = 0.1$).	105
5.3	SDRL values for MHWMA charts ($w = 0.1$).	105
5.4	ARL comparisons for $p = 2$	107

5.5	ARL comparisons for $p = 3$	107
5.6	ARL comparisons for $p = 4$	108
5.7	ARL comparisons for $p = 5$	108
5.8	ARL comparisons for $p = 10$	109
5.9	ARL comparisons for $p = 20$	109
5.10	The simulated dataset.	110
5.11	Simulated bimetal Phase II dataset.	112
7.1	The ARL values of the modified EWMA control charts along with the ARL values of classical EWMA control chart when $C_X = C_Y$, and $n = 5$ for $\rho_{XY} = -0.25, -0.5, -0.75$ and -0.95	165
7.2	The ARL values of the modified EWMA control charts along with the ARL values of classical EWMA control chart when $C_X < C_Y$, and $n = 5$ for $\rho_{XY} = -0.25, -0.5, -0.75$ and -0.95	166
7.3	The ARL values of the modified EWMA control charts along with the ARL values of classical EWMA control chart when $C_X > C_Y$, and $n = 5$ for $\rho_{XY} = -0.25, -0.5, -0.75$ and -0.95	167
7.4	The EQL of the modified EWMA control charts along with the EQL of the classical EWMA control chart for different choices of the size of C_X and C_Y , different values of n and ρ_{XY}	169
7.5	The RARL of the modified EWMA control charts along with the RARL of the classical EWMA control chart for different choices of the size of C_X and C_Y , different values of n and ρ_{XY}	170

8.1	ARL values for the classical EWMA chart with $ARL_0 = 500$	189
8.2	ARL values of the proposed charts when $n = 3$	191
8.3	ARL values of the proposed charts when $n = 10$	192
8.4	SDRL values of the proposed charts when $n = 3$	193
8.5	SDRL values of the proposed charts when $n = 10$	194
8.6	ARL values of the M_X EWMA chart for $n = 3$	196
8.7	ARL values of the M_X EWMA chart for $n = 10$	197
8.8	EQL values of the control charts.	198
8.9	RARL values of the control charts.	198
9.1	ARL and <i>SDRL</i> values of the AHWMA chart when the correlation between the variables is $\rho = 0.05$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w	208
9.2	ARL and <i>SDRL</i> values of the AHWMA chart when the correlation between the variables is $\rho = 0.25$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w	209
9.3	ARL and <i>SDRL</i> values of the AHWMA chart when the correlation between the variables is $\rho = 0.5$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w	209
9.4	ARL and <i>SDRL</i> values of the AHWMA chart when the correlation between the variables is $\rho = 0.75$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w	210

9.5	ARL and $SDRL$ values of the AHWMA chart when the correlation between the variables is $\rho = 0.95$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w	210
9.6	ARL comparisons of the charts.	212
9.7	ARL_0 with bivariate t -distribution	214
9.8	ARL_0 with bivariate gamma distribution	215
9.9	ARL_1 with bivariate t and gamma Distributions	217
9.10	Calculation of the AHWMA chart statistic and its limits	221

List of Figures

1.1	Process variation plot from the two types of causes	2
2.1	The MCUSUM control chart of the Bimetal II data set. The estimated covariance matrices are obtained from the Bimetal I data set. The reference parameter used is $k = 0.5$ and $p = 5$. The upper control limit h , for the known parameters case (Table 2.1) and the corrected upper control limit h_{c1} ; (Table 2.7) used here are the ones that give an in-control ARL of 200.	35
2.2	The MCI control chart of the Bimetal II data set. The estimated covariance matrices are obtained from the Bimetal I data set. The reference parameter used is $k = 0.5$ and $p = 5$. The upper control limit h , for the known parameters case (Table 2.1) and the corrected upper control limit h_{c1} (Table 2.7) used here are the ones that give an in-control ARL of 200.	36
3.1	In-control <i>AARL</i> plot from identity covariance matrix for $r = 0.05$, and different values of p , and m	55
3.2	In-control <i>AARL</i> plot from the AR(1) covariance structure for $r = 0.05$, and different values of p , and m	56

3.3	The MEWMA control charts of the Bimetal II data set. The estimated covariance matrices are obtained from the Bimetal I data set. The smoothing parameter used is $r = 0.05$ and $p = 5$. The upper control limit h , for the known parameters case; given in Table 3.2, and the corrected upper control limit h_c ; given in Table 3.5, used, are the one that give $AARL_0$ of 200.	65
3.4	The scree plot of the gene expression data set	67
3.5	The MEWMA control charts of the gene expression data set. The estimated covariance matrices are obtained from the gene expression Phase I. The smoothing parameter used is $r = 0.05$ and the number of PC axes $k = 55$. The logarithms of upper control limit h , and the corrected upper control limit h_c are shown.	68
3.6	The Shrinkage-based MEWMA control chart of the gene expression for both Phase I and Phase II data set. The estimated shrinkage covariance matrices are obtained from the gene expression Phase I data set. The smoothing parameter used is $r = 0.05$ and $p = 271$. The logarithms of upper control limit h , and the corrected upper control limit h_c are shown.	69
4.1	PTS comparison of the Multivariate CV charts in presence of localized CV disturbances when $n = 10$, $m = 30$ and $\alpha = 0.01$ at different levels of m_1	86
4.2	PTS comparison of Multivariate CV charts in presence of diffuse CV disturbances when $n = 10$, $m = 30$ and $\alpha = 0.01$ at different levels of b	86
4.3	PTS comparison for the Multivariate VN chart when $m = 30$ and $a\alpha = 0.01$ at different levels of n	87

4.4	Scatter plot of the square multivariate CV statistics $\hat{\gamma}_{RR}$, $\hat{\gamma}_{VN}$, and $\hat{\gamma}_{AZ}$ against $\hat{\gamma}^2$ for the carbon fiber tubing data.	89
4.5	Control Chart plots for the three Multivariate CV charts for the carbon fiber tubing data, showing in-control state of the process.	90
4.6	Control Chart plots for the three Multivariate CV charts in presence of localized CV disturbances for the carbon fiber tubing data.	91
5.1	Plot of the logarithms of the ARL values given in Table 5.1.	104
5.2	Plots of the memory-type charts of the simulated dataset	111
5.3	Plots of the memory-type charts of the bimetal dataset.	113
6.1	Two observations may be equidistant from the centroid - this does not mean they will both lie within (or outside) the baseline data cloud.	119
6.2	Probability plot of the different methods from identity covariance matrix	137
6.3	Probability plot of the different methods from the ar(1) covariance matrix	138
6.4	Simulation results for the ecological dataset	139
6.5	Study area and delineation of five sub-areas on the Scotian Shelf off Nova Scotia, Canada	140
6.6	Results of the Area of the Scotian Shelf. The figure includes the non-metric MDS 2D plot, the number of dimension retained at any time t^* , the control chart for the area based on the distance to centroid, and the proposed chart together in a figure.	141
7.1	The estimated standard error plots of the product estimator from different sampling schemes when $C_X = C_Y$	151

7.2	The estimated standard error plots of the product estimator from different sampling schemes when $C_X < C_Y$	153
7.3	The estimated standard error plots of the product estimator from different sampling schemes when $C_X > C_Y$	154
7.4	The classical EWMA control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	171
7.5	The proposed EWMA _P control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	172
7.6	The proposed EWMA _{PRX} control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	172
7.7	The proposed EWMA _{PMX} control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	173

7.8	The proposed EWMA _{P_{RY}} control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	173
7.9	The proposed EWMA _{P_{MY}} control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	174
7.10	The classical EWMA control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$. .	176
7.11	The proposed EWMA _P control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.	176
7.12	The proposed EWMA _{P_{RX}} control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	177
7.13	The proposed EWMA _{P_{MX}} control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	177
7.14	The proposed EWMA _{P_{RY}} control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	178

7.15	The proposed EWMA _{PMY} control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$	178
8.1	The estimated standard error plots of the regression estimator from the case I. . . .	188
8.2	The estimated standard error plots of the regression estimator from the case II. . . .	189
8.3	The plots of the proposed S _X EWMA and T _X EWMA charts for different values of n. . . .	195
8.4	The Classical EWMA and the M _X EWMA control charts CSTR dataset.	199
8.5	The S _X EWMA and T _X EWMA control charts for the CSTR dataset.	200
9.1	Application of the EWMA, HWMA, M _X EWMA, and AHWMA charts.	222

Chapter 1

Introduction

A common motivation for the collection of data is to monitor a system for a change in state. Natural and human systems and measurements are usually subject to variation, which imbues uncertainty into inferences regarding changes in state and limits our ability to determine when a system has fundamentally altered. For example, manufacturing or production processes generally undergo a natural or assignable cause of variations (Abbasi et al., 2015). The natural variation is an intrinsic component of the production process and is usually unavoidable. This type of variation occurs as a results of normal process variation, and can be addressed by process improvement (Montgomery, 2009). In contrast, the assignable causes are not part of the process or do not affect every occurrence of the process but are induced by specific and identifiable causes. Statistical process control (SPC) is a set of methods used to detect (assignable) changes in a process. Among these methods, control charts are the most popular and sophisticated tools for tracking a process and identifying whether it is in control by monitoring essential quality characteristics of interest. Control-chart methods are particularly useful to detect as early as possible, the occurrence of an assignable variation in the production process so that the process can be examined and remedial

actions can be initiated in order to minimise the number of nonconforming products (Montgomery, 2009).

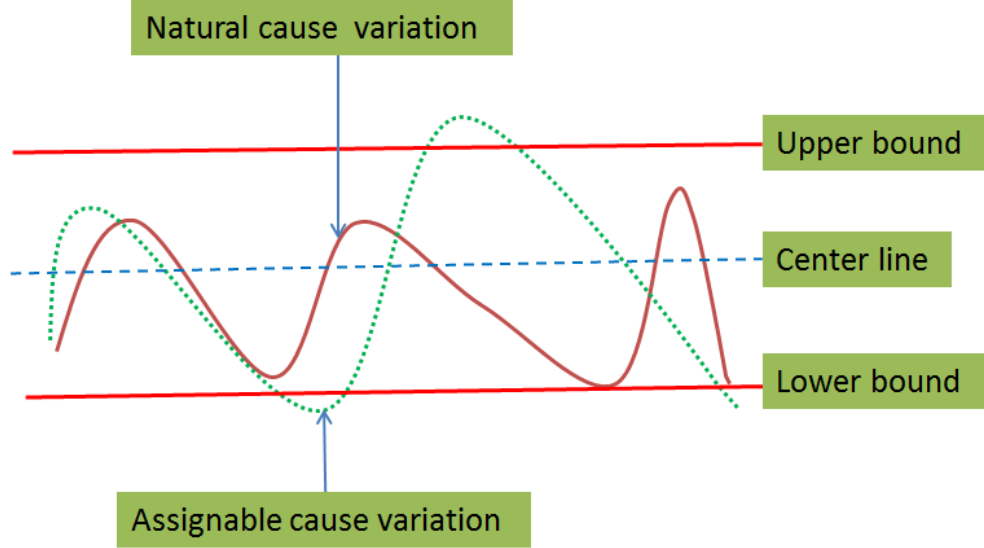


Figure 1.1: Process variation plot from the two types of causes

Control charts are classified into (i) memory-less control charts and (ii) memory-type control charts; these are useful for monitoring large and small-to-moderate shifts in the process, respectively. For example, the Shewhart chart is a memory-less type control chart for a univariate variable that uses only the current process information without referring to any past behavior of the process. It is very effective for detecting a large shift in the process mean (i.e., $\delta \geq 2$, where δ is the size of the shift in standard deviation units (Testik et al., 2003)). The homogeneously weighted moving average (HWMA) control chart by Abbas (2018), is a memory-type chart proposed for efficient monitoring of small (i.e., $\delta \leq 0.5$) to moderate (i.e., $0.5 < \delta < 2$) shifts in the process mean. Other memory-type charts include the EWMA chart by Roberts (1959), the CUSUM chart by Page (1961), the mixed EWMA-CUSUM chart proposed by Abbas et al. (2013) and their multivariate extensions. These charts are capable of signalling when a process is con-

sidered in control vs out of control (Jackson, 1991), and are applied in two monitoring phases (Montgomery, 2009). In Phase I, a historical reference sample is analysed to establish the values and stability of process parameters while in the in-control state. If the in-control parameter values are unknown, the data from Phase I are used to estimate these values and their control limits (Abbasi and Adegoke, 2018). In Phase II, the process parameters are monitored and checked for departure from the in-control state. While Phase II values (or statistics) remain inside the Phase I limits, the process is considered to be in control; if they go outside the control limits, this indicates that the process may be out-of-control and remedial actions are triggered.

Univariate control charts are widely used in most of today's industries for efficient monitoring of a process characteristic (or variable). They have the advantages of being relatively simple to implement and interpret, and can promptly detect process shifts. Several applications of classical univariate control charts focus on monitoring the process in situations where the process variable is independent of other variables. For examples, see Abtew et al. (2018); El-Din et al. (2006); Hayes et al. (1997); Grigg, Nigel P and Daly, Jeannette and Stewart (1998); Benneyan (1998a,b); Srikaeo and Hourigan (2002); De Vries and Conlin (2003); De Vries and Reneau (2010); Madsen and Kristensen (2005) for a variety of applications of classical univariate control charts. In many cases, direct measurements of the process of interest may be measured alongside some auxiliary variables with which it is correlated. Where available, auxiliary variables may be used to improve the efficiency of the classical univariate control chart schemes, using standard methods such as regression (Riaz, 2008b). The concept of using supplemental information to provide a more efficient estimate of a population parameter is popular in the field of survey sampling (Cochran, 1977). Several researchers have studied and recommended the introduction of auxiliary variables into

the monitoring scheme of a process variable of interest, and have proposed a variety of different control charts tools for this purpose. For example, Abbas et al. (2014a) proposed an exponentially weighted moving average (EWMA)-type control chart where the process mean is regressed on the auxiliary variable. This method performed better than either the classical univariate or bivariate EWMA control charts, particularly for detecting small-to-moderate shifts in the process mean.

In many applications, several variables are measured at each time point. Assessing each variable independently can be inefficient and misleading; a particular observation may not look unusual when examined on its own, but may be unusual when considering the joint multivariate structure of the data. As a result, multivariate control charts (MCC) methods, in which several related process parameters are jointly monitored (Seif et al., 2011), comprise a rapidly developing area of research in SPC. Several MCC tools that use the relationships among the variables to provide efficient monitoring schemes for identifying any changes in the quality of products have been proposed (e.g., Hotelling (1947); Crosier (1988); Pignatiello and Runger (1990); Lowry et al. (1992)).

When the underlying process parameters of interest (e.g., the mean vector and covariance matrix), used in the development of the MCC methods are unknown, they need to be estimated from the historical Phase I dataset (see for example, Hotelling (1947), Crosier (1988), Pignatiello and Runger (1990), and Lowry et al. (1992)). One of the most significant modern problems in the use of MCC tools is the estimation of the covariance matrix. While the empirical estimate of the covariance matrix provides an unbiased and consistent estimator for a low-dimensional data matrix with an adequate Phase I sample, its inconsistent performance in the high-dimensional framework is well known (Melorose et al., 2015). In circumstances where the number of variables p approaches or exceeds the number of Phase I samples m , the performance of charts based on the empirical

covariance estimate has been shown to be poor (Stein and Others, 1956; Schäfer and Strimmer, 2005; Giovannelli and Idier, 2015). It suffers from a distortion of its eigenstructure (Giovannelli and Idier, 2015); specifically, the empirical covariance estimate inflates the higher eigenvalues and deflates the lower eigenvalues, rendering it ill-conditioned, unless the sample m is much larger than the dimension p . A standard solution to the problem of unstable estimation of the covariance matrix from a high-dimensional data set is to regularize the estimate (Melorose et al., 2015). One such method of regularization is to adjust the eigenvalues in a way that guarantees a positive definite covariance matrix (Higham, 1988). Another option is to shrink the empirical covariance matrix towards a target matrix (Stein and Others, 1956; Schäfer and Strimmer, 2005). There are also several methods based on Gaussian graphical models (i.e., penalized likelihood and graphical lasso) for estimating the inverse of high-dimensional covariance matrix, developed in genetics and Bayesian networks, among other fields (Whittaker, 2009).

Most control charts focus on monitoring either the process mean or the process variability, and they require the process mean to be stable and independent of the process standard deviation (Yeong et al., 2016). However, in many real-life processes, the process standard deviation is dependent on the mean, and the mean is not constant. In such cases, it is more appropriate to monitor the coefficient of variation (CV). Several methods for monitoring multivariate CV have been proposed in the control chart literature. However, the proposed multivariate CV control-chart statistic, (γ) , is a function of the process parameters, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. In practice, these parameters are rarely known and, therefore, must be estimated from samples taken when the process is assumed to be in control. The Phase II performance of the multivariate CV depends on the stability of the estimated parameters obtained from Phase I. Optimal performance requires any changes in these

parameters to be detected as early as possible (Riaz et al., 2011).

Control charts have proven very useful in monitoring industrial and manufacturing processes consisting of univariate or independent (approximately) multivariate-normal data (Montgomery, 2009). The statistical properties of traditional control charts are maintained only if these assumptions are satisfied. However, in some cases, the underlying process is not normal and, as a result, the traditional charts can be inefficient (Chakraborti et al., 2001). The necessity of this assumption has largely restricted the application of traditional control charts to industrial process monitoring (Tuerhong and Bum Kim, 2014). For example, ecological monitoring often provides data in the form of abundance counts of multiple species, which fail to abide by the assumptions of multivariate normality. The extension of control chart methods to a broader range of contexts, such as ecological monitoring, requires more robust methods that are capable of dealing with a wider range of multivariate data.

1.1 Thesis Contribution

This thesis develops control chart methods for efficient monitoring in Phase I and Phase II, and has two principal aims:

- to develop efficient new parametric and non-parametric multivariate control-chart methods for detecting changes in the process parameters of interest (Chapters 2-6).
- to develop more efficient univariate control chart methods for monitoring small shifts in the process mean when the process variable is observed along with a correlated auxiliary variable (Chapters 7-8).

1.2 Thesis Outline

This thesis is organised into ten self-contained chapters. We outline the content of the chapters that comprise this thesis, including the control-chart methods examined and the specific novelty attending the studies in each chapter below (Table 1.1).

Table 1.1: The study done including the control chart methods examined, the novelty of each chapter and the size of the subgroup sample (n) considered. Some of the control chart methods examined include multivariate cumulative sum (MCUSUM), exponentially weighted moving average (EWMA), multivariate EWMA (MEWMA), homogeneously weighted moving average (HWMA), and multivariate HWMA (MHWMA).

Chapters	Study	Control charts	n	Novelty
Chapter 2	On the effect of parameters estimation	MSCUSUM charts	1	Investigate and propose shrinkage estimates of covariance
Chapter 3	On the effect of parameters estimation	MEWMA chart	1	Investigate and propose shrinkage estimates of covariance
Chapter 4	On the coefficient of variation (cv) charts in Phase I	Multivariate CV charts	$n > 1$	Suggest appropriate multivariate CV chart
Chapter 5	A new multivariate chart for monitoring small shifts	MHWMA chart	all n	Propose an efficient chart for monitoring small shift in the process mean vector
Chapter 6	Contribution to non-normality monitoring	A changed-point multivariate chart	all n	Implement efficient tool for ecological data monitoring
Chapter 7	Contribution to control chart for negatively correlated bivariate dataset under ranked sampling schemes	EWMA chart	> 1	Implement EWMA chart under different sampling schemes for negatively correlated bivariate dataset
Chapter 8	Contribution to control chart for correlated bivariate dataset under ranked sampling schemes	EWMA chart	> 1	Implement Ranked-based EWMA chart for correlated bivariate dataset
Chapter 9	Efficient monitoring of the process mean using an auxiliary variable	HWMA chart	all n	Propose an efficient chart for monitoring small shift of the process mean using an auxiliary variable

Chapter 2 describes an improvement to the estimation procedure used for μ_0 and Σ_0 in the development of the MCUSUM control chart, and the multivariate CUSUM #1 (MCI) control chart proposed by Pignatiello and Runger (1990), for individual monitoring. These multivariate CUSUM charts have been shown to have better run-length performance than other multivariate CUSUM charts (Woodall and Mahmoud, 2005). Here, we propose the use of a shrinkage estimate of Σ_0 in Phase I, and compare the resultant MCUSUM and MCI control charts with those obtained using some other methods of estimating Σ_0 . The shrinkage estimate is a weighted combination of the empirical estimate and a target matrix. The findings from this study will provide practitioners with a more efficient method of applying the MCUSUM and MCI charts for individual-observation monitoring when Σ_0 is unknown.

Chapter 3 studies the relative performance of the MEWMA control chart from the empirical estimate of the $\boldsymbol{\mu}_0$ vector, and three different estimates of $\boldsymbol{\Sigma}_0$ from Phase I data; namely: (i) the classical empirical estimate, (ii) the MSSD, and (iii) a shrinkage estimate of $\boldsymbol{\Sigma}_0$ proposed here. This study provides control-chart practitioners with an appropriate choice of the estimated parameters to be used in the MEWMA control charts for individual-observation monitoring, and also with appropriate estimates of $\boldsymbol{\Sigma}_0$ when large Phase I datasets are not available.

Chapter 4 provides a comprehensive study of the Phase I analysis for multivariate CV control charts. We investigated the performance of a variety of multivariate CV charts in Phase I, considering both diffuse symmetric and localized CV disturbance scenarios, and using probability to signal (PTS) as a performance measure. The PTS performance measure was found to yield a useful application in Phase I analysis for both location and dispersion parameters to yield an efficient control chart for monitoring multivariate CV.

Chapter 5 proposes a new memory-type multivariate charting procedure, namely, the multivariate homogeneously weighted moving average (MHWMA) control chart. Like other memory-type charts, MHWMA uses the current observation and past observations. However, previous methods allocated equal weight across all observations, including the current one. With our proposed MHWMA method, the weight of the current observation can be specified separately, with the remaining weight then being allocated equally across previous observations. We showed that this approach can provide more efficient monitoring of small shifts in the process mean vector compared to other memory-type multivariate charting procedures.

Chapter 6 proposes a new multivariate control chart for monitoring multivariate ecological data. The proposed chart is based on a change-point method where the currently observed com-

munity is evaluated based on all of the previously observed communities at a given site, and does not require prior knowledge of the ecosystem's behavior before the monitoring begins. The method can take the multivariate dataset or the dissimilarities between every possible pairs of samples as input. When dissimilarities between every possible pair of the samples is available, the method first represents these sample distances as points in a principal co-ordinates space. Then, we use these co-ordinates to monitor and identify, as quickly as possible, any change in the mean vector of the multivariate ecological dataset. The method does not require any parametric model assumptions, and it can be based on any dissimilarity measure of choice. A permutation procedure is employed to obtain the control-chart limits for a suitable distance-based model through time.

Chapter 7 proposes an improvement to the performance of the classical exponentially weighted moving average (EWMA) control chart, by making use of auxiliary information in the form of a variable that is negatively correlated with the process variable of interest. The charts are developed using different sampling schemes: simple random sampling, ranked set sampling (RSS) and median ranked set sampling (MRSS), and we evaluate their performance using average run length, and other performance measures such as extra quadratic loss and relative average run length.

Chapter 8 proposes enhancements to the applications of the EWMA control chart for those scenarios where the exact measurement of process units is difficult or expensive, but the visual ordering of the units can be done easily. The proposed charts use an auxiliary variable that is (positive or negatively) correlated with the process variable to provide efficient monitoring of shifts in the process mean and are formulated based on ranked set sampling (RSS) and median RSS schemes (MRSS). The proposed chart is compared with some existing charts.

Chapter 9 presents an efficient control chart method for monitoring small shifts in the process

mean for scenarios where the process variable is observed along with a correlated auxiliary variable. The proposed chart, called an AHWMA chart, is a homogeneously weighted moving average type control chart that uses information from both the process variable and auxiliary variable in the form of a regression estimator to provide an efficient and unbiased estimate of the mean of the process variable. We provide the design structure of the chart and examine its performance in terms of its run length properties. We provide a detailed study of the chart's robustness to non-normal distributions, and give some recommendations on the application of the chart when the process parameters are unknown.

Chapter 10 presents a summary of the main findings and points to future research directions to pursue following this thesis work.

For each chapter, we used both simulated and real-life datasets to test and compare methods. All studies were done using the *R* language for statistical computing (R Core Team, 2013).

Chapter 2

Multivariate cumulative sum control charts for individual-observations monitoring when parameters are estimated

”This is the peer reviewed version of the following article: “Adegoke, N. A., Smith, A. N., Anderson, M. J., Abbasi, S. A., & Pawley, M. D. (2018). Shrinkage estimates of covariance matrices to improve the performance of multivariate cumulative sum control charts. *Computers & Industrial Engineering*, 117, 207-216”, which has been published in final form at <https://doi.org/10.1016/j.cie.2018.02.008>. This article may be used for non-commercial purposes in accordance with Elsevier Terms and Conditions for Use of Self-Archived Versions.”

Multivariate cumulative sum control charts require knowledge of the in-control process covariance parameters. Here, we show that the performance of the multivariate cumulative sum control charts for individual-observation monitoring is affected by the estimation of parameters unless the Phase I sample size is large. When only a small Phase I sample size is available, we propose the use of a shrinkage estimate. The average run length performance of multivariate cumulative sum control charts obtained using the shrinkage estimate is superior to the other methods examined in this study. The improved performance of the control charts using the shrinkage estimate is also demonstrated via an illustrative case study of Bimetal data, in which measurements of four properties of bimetal brass and steel thermostats are monitored, and a shift in the multivariate centroid is detected earlier using the shrinkage-based method.

2.1 Introduction

The cumulative sum (CUSUM) chart, proposed by Page (1954), is a control chart used for detecting small to moderate changes in the process measurements or the means of the process values measured from a manufacturing or industrial process. The CUSUM control chart is an efficient tool in process monitoring for detecting any changes that may affect the quality of a product, and serves as an alternative control chart to the Shewhart control chart (Shewhart, 1931), which is more useful for detecting large shifts. The CUSUM control chart is particularly useful in the context of normally distributed process data (Duncan, 1974), and involves plotting the cumulative total against the observation number (Johnson, 1961). Several authors have contributed to CUSUM chart methodology, including Ewan and Kemp (1960), Brook and Evans (1972), Lucas and Crosier (1982), and Woodall (1983), among others.

Crosier (1988) initially proposed the multivariate cumulative sum control chart (MCUSUM), since then, several versions of the MCUSUM charts have been suggested in the literature of Statistical Process Control (SPC). Among others, Pignatiello and Runger (1990) and Ngai and Zhang (2001) recommended different multivariate CUSUM charts that are effective for detecting small to moderate shifts in either process measurements or the process mean vector. Also, because the MCUSUM control chart and its modifications combine and use information from several samples, they are particularly effective and efficient for detecting shifts in a process with individual-observation monitoring.

In most of the practical applications of the MCUSUM chart and its modifications, it is assumed that the in-control parameters are known from Phase I, and the in-control limits are calculated from the assumed known in-control parameters (Crosier, 1988). The control charts are then applied in Phase II to check whether future observed values are within the in-control limits. If successive observed future values fall within the in-control limits, the process is believed to be in control. Observed future values falling outside the control limits indicate that the process may be out-of-control, and remedial responses are then required (Montgomery, 2009).

Thus, the Phase II charting schemes are based on the requirement that the process parameters are known from Phase I. For example, when multivariate data are drawn from the multivariate normal distribution, the in-control process parameters: mean vector ($\boldsymbol{\mu}_0$), and covariance matrix ($\boldsymbol{\Sigma}_0$), are assumed to be known, which simplifies the Phase II development of the control charts. However, the process parameters are generally unknown, and control limits for Phase II are usually based on parameters that are estimated from Phase I data. In the situation where $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ are unknown, the empirical estimate of $\boldsymbol{\mu}_0$ of the p -variate random variables or process data \mathbf{Y} ,

is obtained by averaging the pooled sample obtained during Phase I. An empirical estimate of Σ_0 from the pooled Phase I samples is commonly also obtained using unbiased variance and covariance estimators.

The problem of using the empirical estimate of Σ_0 for the Hotelling's T^2 control chart is that it leads to a chart that may not be efficient in detecting a shift in μ_0 , because Σ_0 may be poorly estimated (Sullivan and Woodall, 1996). Hence, Sullivan and Woodall (1996) proposed a method of estimation that uses the differences between successive observations (hereafter called SW). Also, Holmes and Mergen (1993) proposed the use of a mean square successive difference (hereafter called MSSD) approach for estimating Σ_0 . There is almost no work on the performance of the multivariate CUSUM control charts for individual-observation data, the situation where these charts have found greatest application.

In this study, we propose an improvement to the estimation procedure used for μ_0 and Σ_0 in the development of the MCUSUM control chart, and the multivariate CUSUM #1 (MCI) control chart proposed by Pignatiello and Runger (1990), for individual monitoring. These multivariate CUSUM charts have been shown to have better run-length performance than other multivariate CUSUM charts (Woodall and Mahmoud, 2005). Here, we propose the use of a shrinkage estimate of Σ_0 in Phase I, and compare the resultant MCUSUM and MCI control charts with those obtained using other methods of estimating Σ_0 , including: the empirical unbiased estimate, the MSSD, and the SW. The shrinkage estimate is a weighted combination of the empirical estimate and a target matrix. The results from this study will provide practitioners with a more efficient method of applying MCUSUM and MCI control charts for individual-observation data when Σ_0 is unknown.

The rest of the chapter is organized as follows: In Section 2.2, we describe multivariate cu-

mulative sum control charts; specifically, the MCUSUM and MCI control charts. In Section 2.3, we give the different proposed estimates of Σ_0 , including a shrinkage estimate. Section 2.5 shows the effects of the different estimators on the in-control average run-length (ARL) performance of the MCUSUM and MCI control charts for individual monitoring. Adjusted control limits for the MCUSUM and MCI control charts from individual-observation monitoring with estimated parameters are given in Section 2.6, followed by simulation procedures to assess the out-of-control run-length performance of the corrected limits in Section 2.7. We present the estimated corrected limits of the charts using a shrinkage estimate of the covariance matrix and the empirical estimate of the location parameter in Section 2.8. In Section 2.9, the procedure is applied to a Bimetal data set, with a concluding discussion in Section 2.10.

2.2 Multivariate cumulative sum control charts

The cumulative sum (CUSUM) statistic for a univariate random variable y with mean μ_0 , is given as the cumulative sum of the deviation of the sample value from the target value (Page, 1961).

The chart is obtained by plotting the CUSUM statistic

$$S_i = \sum_{j=1}^i (y_j - \mu_0) \quad (2.1)$$

against the sample number i , where μ_0 is the true process mean and S_i is the cumulative sum up to and including the i th sample. The chart is more effective when the sample size $n = 1$, and is often used in chemical and process industries where subgroup sizes are indeed usually of size $n = 1$ (Montgomery, 2009). When $n > 1$, the plotting statistic in Equation 2.1 can be obtained

by replacing the sample value y by the average of the j th sample (i.e., \bar{y}_j) The one-sided upper and lower control limits for $n = 1$ subgroup size are given as:

$$S_i^+ = \max(0, y_i - (\mu_0 + k) + S_{i-1}^+) \quad (2.2)$$

$$S_i^- = \min(0, y_i - (\mu_0 - k) + S_{i-1}^-) \quad (2.3)$$

where $S_0^+ = S_0^- = 0$, $k = \frac{\delta\sigma}{2}$ is the reference value, σ is the process standard deviation and δ is the size of the shift in the process mean. If either the S_i^+ or S_i^- exceeds the decision interval h , the process is deemed to be out of control. The value of h is chosen for a fixed performance measure of the chart.

When a p -variate dataset is monitored, the p monitored quality characteristics can be represented by a $p \times 1$ random vector \mathbf{Y} . We assume that the random vector \mathbf{Y} is independent and identically distributed as multivariate normal with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. Two multivariate CUSUM charts were proposed by Crosier (1988). The one with better ARL performance obtains the CUSUM vector directly from the multivariate observation, and the MCUSUM vector for the observed vector \mathbf{y}_i is given as:

$$C_i = [(\mathbf{S}_{i-1} + \mathbf{y}_i - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\mathbf{S}_{i-1} + \mathbf{y}_i - \boldsymbol{\mu}_0)]^{1/2}, \quad (2.4)$$

where

$$\mathbf{S}_i = \mathbf{0} \quad \text{if } C_i \leq k$$

$$\mathbf{S}_i = (\mathbf{S}_{i-1} + \mathbf{y}_i - \boldsymbol{\mu}_0)(1 - k/C_i) \quad \text{if } C_i > k,$$

$\mathbf{S}_0 = \mathbf{0}$ and $k > 0$. The MCUSUM control chart signals when $T_i^2 = [\mathbf{S}_i' \boldsymbol{\Sigma}_0^{-1} \mathbf{S}_i] > h_1$. The parameters of the chart k and $h_1 (> 0)$, are chosen to give the desired in-control ARL performance of the chart (Crosier (1988)).

Two directionally invariant multivariate CUSUM charts were proposed by Pignatiello and Runger (1990); the one with better ARL performance is the MCI chart. Here, the CUSUM vector for the observed vector \mathbf{y}_i is given as:

$$\mathbf{C}_i = \sum_{j=i-n_i+1}^i (\mathbf{y}_j - \boldsymbol{\mu}_0). \quad (2.5)$$

$$T_i = \max \left\{ \sqrt{\mathbf{C}_i' \boldsymbol{\Sigma}_0^{-1} \mathbf{C}_i} - kn_i, 0 \right\},$$

and

$$n_i = \begin{cases} n_{i-1} + 1, & \text{if } T_{i-1} > 0 \\ 1, & \text{if otherwise} \end{cases}$$

where n_i is interpreted as the number of subgroups up to the most recent CUSUM statistic. The MCI control chart signals when $T_i > h_2$, for $h_2 > 0$ and $k > 0$. The values of k and h_2 are chosen to give the desired in-control ARL performance for the chart. Because the MCUSUM and MCI charts are directionally invariant charts, their ARL performances depend on $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ only through the non-centrality parameter given as

$$\delta = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)}. \quad (2.6)$$

It is recommended to choose the value of $k = \frac{\delta}{2}$. We refer the reader to the original works by Crosier

(1988), and Pignatiello and Runger (1990) for more comprehensive details on the development of these charts.

2.3 Estimates of the mean vector and covariance matrix

The empirical estimate of $\boldsymbol{\mu}_0$ of the p -variate random variables or process data \mathbf{Y} , is obtained by taking the average of the pooled sample in Phase I, given by

$$\bar{y}_k = \frac{1}{m} \sum_{i=1}^m y_{ik}$$

where $k = 1, \dots, p$, p is the number of quality characteristics and m is the Phase I sample size.

Thus, $\bar{\mathbf{y}} = (\bar{y}_1, \bar{y}_2, \dots, \bar{y}_p)'$ is a $1 \times p$ vector of means of the p -dimensional variables.

In this case, there is no subgroup information to calculate an estimate of $\boldsymbol{\Sigma}_0$. A common approach in practice is to obtain an empirical estimate of $\boldsymbol{\Sigma}_0$ from the m Phase I samples as:

$$\mathbf{S} = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})', \quad (2.7)$$

where the main diagonal of \mathbf{S} contains the sample variances $s_k^2 = \frac{1}{m-1} \sum_{i=1}^m (y_{ik} - \bar{y}_k)^2$, and the off-diagonal of \mathbf{S} contains the sample covariances $s_{kk'} = \frac{1}{m-1} \sum_{i=1}^m (y_{ik} - \bar{y}_k)(y_{ik'} - \bar{y}_{k'})$ (Montgomery, 2009). The proposed estimator of $\boldsymbol{\Sigma}_0$ by Sullivan and Woodall (1996) is

$$\hat{\Sigma}_{SW} = \frac{\sum_{j=1}^m \mathbf{F}_j}{m-1}, \quad (2.8)$$

where

$$\mathbf{F}_i = \mathbf{F}_{i-1} + (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})' \quad \text{where, } \mathbf{F}_0 = 0, \quad i = 1, 2, \dots, m$$

To make the Hotelling's T^2 control chart more sensitive to process changes, Holmes and Mergen (1993) proposed the use of the mean square successive difference (MSSD) approach for estimating Σ_0 . They used the difference among consecutive observations as follows:

$$\mathbf{u}_i = \mathbf{y}_{i+1} - \mathbf{y}_i, i = 1, 2, \dots, m - 1 \quad (2.9)$$

and the MSSD estimator of Σ_0 is

$$\hat{\Sigma}_{MSSD} = \frac{\mathbf{U}'\mathbf{U}}{2(m-1)}, \quad (2.10)$$

where \mathbf{U} is a $(m-1) \times p$ matrix of $\mathbf{u}_i, i = 1, 2, \dots, m-1$.

2.4 Shrinkage estimate of the covariance matrix

The shrinkage estimator is a weighted average of the empirical unbiased estimate \mathbf{S} given in Equation (2.7), and a target matrix \mathbf{T} , and is given as:

$$\hat{\Sigma}_S = \lambda \mathbf{T} + (1 - \lambda) \mathbf{S}, \quad (2.11)$$

where, $\lambda \in [0, 1]$ denotes the shrinkage intensity. $\lambda = 0$ implies $\hat{\Sigma}_S = \mathbf{S}$, and $\lambda = 1$ gives $\hat{\Sigma}_S = \mathbf{T}$.

The optimal value of the shrinkage intensity, λ , is obtained by minimizing the risk function:

$$R(\lambda) = \|\hat{\Sigma}_S - \Sigma_0\|_F^2, \quad (2.12)$$

where, $\|\cdot\|_F^2$ is the squared Frobenius norm, which is a quadratic form associated with the inner product. Analytically, the optimal shrinkage intensity is given as:

$$\lambda^* = \frac{\sum_{k=1}^p \sum_{k'=1}^p \text{Var}(s_{kk'}) - \text{Cov}(t_{kk'}, s_{kk'})}{\sum_{k=1}^p \sum_{k'=1}^p E(t_{kk'} - s_{kk'})^2}, \quad (2.13)$$

where $t_{kk'}$ and $s_{kk'}$ are the observed values of \mathbf{T} and \mathbf{S} , respectively. This method of estimating the optimal shrinkage intensity is shown to be consistent, and the shrinkage estimator for Σ_0 is more accurate than the empirical sample covariance matrix, even for small m and p (Ledoit and Wolf, 2004a). Further, the shrinkage estimator is better conditioned than the empirical estimate of the covariance matrix, and asymptotic properties hold well even with finite samples (Warton, 2008).

We followed the approach of Schäfer and Strimmer (2005) and replaced the quantities in Equation (2.13) by their unbiased estimates. This gives an estimated optimal shrinkage intensity of the form:

$$\hat{\lambda}^* = \frac{\sum_{k=1}^p \sum_{k'=1}^p \widehat{\text{Var}}(s_{kk'}) - \widehat{\text{Cov}}(t_{kk'}, s_{kk'})}{\sum_{k=1}^p \sum_{k'=1}^p E(t_{kk'} - s_{kk'})^2}. \quad (2.14)$$

The estimated optimal shrinkage intensity of the diagonal covariance target \mathbf{T} and the sample

covariance matrix \mathbf{S} is given as:

$$\hat{\lambda}^* = \frac{\sum_{k \neq k'} \widehat{Var}(s_{kk'})}{\sum_{k \neq k'} s_{kk'}^2}, \quad (2.15)$$

where the target covariance \mathbf{T} with elements

$$t_{kk'} = \begin{cases} s_{kk} & \text{if } k = k' \\ 0 & \text{if } k \neq k' \end{cases} \quad (2.16)$$

is a diagonal matrix with the unbiased sample variances of the p quality characteristics along the main diagonal, (i.e., $\mathbf{T} = \text{diag}(\mathbf{S})$), so that shrinkage is only applied to the off-diagonal elements (Schäfer and Strimmer, 2005).

2.5 Assessing the in-control performance of MCUSUM and MCI control charts using simulation

2.5.1 Simulation method

We compared the in-control performance of individual-observation MCUSUM and MCI charts based on the four methods of estimating Σ_0 given by equations (2.7), (2.8), (2.10) and (2.11), using simulated data.

The following procedures were used in our simulation study:

- Phase I

- (i) Generate m baseline observations of size $n = 1$ from a p -variate normal distribution with mean vector $\boldsymbol{\mu}_0 = \mathbf{0}$ and $\boldsymbol{\Sigma}_0 = \mathbf{I}_p$. Here, p is the number of variables used, and \mathbf{I}_p is a $p \times p$ identity matrix.
 - (ii) Compute both the estimated covariance matrix and the empirical estimate of the location parameter. These estimated parameters were used to set control limits for Phase II.
- Phase II
 - (i) At each time i , generate a random vector from the p -variate normal distribution (denoted as \mathbf{y}_i). In this case, we applied shift of size δ to the mean vector $\boldsymbol{\mu}_0$ ($\delta = 0$ for the in-control state).
 - (ii) Using the estimated parameters from Phase I and \mathbf{y}_i , compute the MCUSUM and MCI control charts test statistic values at each step i , and compare against the corresponding upper control limits given in Table 2.1.
 - (iii) Repeat steps (i-ii) and record the iteration number (run length, RL) that gives the first out-of-control signal.
 - Repeat Phase I and Phase II processes K times; where K is the number of the simulation runs. We used $K = 50,000$.
 - The average of the run lengths across simulations (i.e., ARL_0) was recorded.

Our simulations included number of Phase I samples $m = \{30, 40, 50, 70, 100, 200, 300 \text{ or } 500\}$ (all with $n = 1$), having $p = 2, 3$, or 5 variables. For establishing control limits from Phase I, we used three different values of the reference parameter $k = \{0.25, 0.50 \text{ or } 1.00\}$. Following Mahmoud

and Maravelakis (2013), values of the upper control limits for each simulation were determined such that the charts would produce an ARL_0 of 200 under known μ_0 and Σ_0 (see Table 2.1 for examples for some selected values of p , and section 2.6 for correcting control limits to give ARL_0 of 200 when the population parameters are unknown). Our study differs from that of Mahmoud and Maravelakis (2013) in the way that we have considered methods for estimating Σ_0 when there are no subgroups within time periods. We also give the in-control ARL values of the charts for $p = 5$ and $k = 0.5$ that produce ARL_0 of 500 when parameters are known.

Table 2.1: The upper control limits: h_1 and h_2 , of the MCUSUM and MCI charts, respectively, that produce in-control ARL of 200 when the charts' statistics are obtained from known parameters, for different values of p and k

		k		
<i>Chart</i>	p	0.25	0.50	1.00
MCUSUM	2	8.64	5.49	3.00
	3	10.85	6.90	3.77
	5	14.81	9.40	5.21
MCI	2	7.49	4.78	2.71
	3	8.75	5.52	3.14
	5	10.80	6.81	3.92

2.5.2 Simulation results of the in-control ARL

The in-control ARL of the MCUSUM and MCI charts, for all parameter values, was less than the benchmark of 200 which would be achieved if the known parameters were used to establish the control limits (given in Tables 2.2-2.4), demonstrating that having to estimate the population parameters compromises the performance of these control charts by increasing the frequency of false out-of-control signals. For example, when $m = 30$ and $p = 2$, the in-control ARL values of the MCUSUM control chart ranged from 100.12 to 123.57 among the methods (see Table 2.3). For the MCI control charts, the in-control ARL ranged from 100.55 to 127.42. However, as the value

of m increases the number of false alarms from all methods decreases, and the differences between in-control ARL values from the charts obtained using known parameters and those obtained using estimated parameters get smaller.

Table 2.2: In-control ARL of the MCUSUM and MCI control charts for $p = 2$ or 5, when different Phase I samples (m), of size $n = 1$, are used in estimating the unknown parameters. We used $k = 0.25$, and the upper control limits that give in-control ARL of 200 when the parameters are known.

<i>Chart</i>	p	<i>Method</i>	m								
			30	40	50	70	100	150	200	300	500
<i>MCUSUM</i>	2	<i>Empirical</i>	81.96	94.02	102.21	115.18	129.47	144.46	152.25	163.75	176.24
		<i>SW</i>	82.39	93.81	102.31	115.45	129.38	143.64	153.06	164.05	174.3
		<i>MSSD</i>	81.14	91.47	101.39	114.11	127.23	144	151.91	162.59	173.72
		<i>Shrinkage</i>	93.61	102.67	109.13	122.17	135.19	148.41	155.14	166.22	177.62
	5	<i>Empirical</i>	43.37	52.77	60.53	73.17	88.09	106.23	117.87	134.85	154.72
		<i>SW</i>	43.41	52.5	60.51	73.23	87.93	105.97	118.88	135.9	154.32
		<i>MSSD</i>	38.82	47.93	55.86	68.87	84.15	101.97	114.26	133.08	151.11
		<i>Shrinkage</i>	54.13	63.32	70.93	83.77	96.83	114.19	126.42	141.82	159.23
	2	<i>Empirical</i>	83.5	94.99	103.54	118.24	131.82	145.04	154.29	166.07	176.81
		<i>SW</i>	83.09	94.99	103.3	116.18	132.03	145.83	154.19	165.85	175.82
		<i>MSSD</i>	80.9	91.96	101.36	114.03	129.22	143	152.15	165.84	175.03
		<i>Shrinkage</i>	93.36	104.19	110.56	123.77	136.2	149.92	158.14	167.58	177.74
	5	<i>Empirical</i>	37.4	47.34	55.58	70.02	86.45	106.75	120.02	139	157.18
		<i>SW</i>	37.35	47.17	56.13	69.78	86.43	107.03	121.02	137.53	156.21
		<i>MSSD</i>	32.58	42.14	50.19	64.3	81.91	101.02	116.38	134.36	154.1
		<i>Shrinkage</i>	50.57	59.85	68.36	83.46	99.6	118.18	131	146.53	163.69

Generally, the in-control ARL values depend on the reference parameter k , the number of Phase I samples m , the number of quality characteristics p , and the method used to estimate Σ_0 . Specifically, the charts performed better (i.e., had greater ARL_0) with a greater number of Phase I samples m , higher values of the reference parameter k , and lower dimensionality p . For every parameter combination we considered, the in-control performance was best for our proposed shrinkage method. The empirical estimate and the SW gave similar performance, and the MSSD gave the highest number of false signals. As expected, the in-control ARL performance of the MCI control chart generally performed better than the in-control behavior of the MCUSUM control

Table 2.3: In-control ARL of the MCUSUM and MCI control charts for $p = 2$ or 5 , when different Phase I samples (m), of size $n = 1$, are used in estimating the unknown parameters. We used $k = 0.5$, and the upper control limits that give in-control ARL of 200 when the parameters are known.

<i>Chart</i>	p	<i>Method</i>	m								
			30	40	50	70	100	150	200	300	500
<i>MCUSUM</i>	2	<i>Empirical</i>	100.76	111.73	121.41	134.33	146.38	159.95	167.37	175.69	184.36
		<i>SW</i>	100.98	110.71	121.36	134.3	145.61	159.81	166.83	174.46	182.55
		<i>MSSD</i>	100.12	111.17	119.26	132.89	145.56	159.6	166.07	175.58	184.45
		<i>Shrinkage</i>	123.57	134.39	138.14	147.65	157.87	168.36	172.35	180.31	187.79
	5	<i>Empirical</i>	37.82	48.41	57.24	73.41	90.9	111.99	125.21	144.18	161.54
		<i>SW</i>	37.65	48.84	57.7	73.51	91.3	111.8	123.77	144.09	163.35
		<i>MSSD</i>	31.44	41.39	50.16	64.75	82.9	103.88	118.99	137.82	157.82
		<i>Shrinkage</i>	60.32	71.52	80.82	96.52	112.97	130.3	142.12	158.05	171.5
<i>MCI</i>	2	<i>Empirical</i>	102.64	113.23	122.52	136.94	149.27	162.59	170.68	179.68	187.02
		<i>SW</i>	102.84	112.29	123.24	136.82	150.13	162.29	169.57	177.60	185.94
		<i>MSSD</i>	100.55	111.32	119.64	133.95	145.82	160.82	169.79	178.32	187.30
		<i>Shrinkage</i>	127.42	136.45	143.37	151.97	162.37	172.81	176.03	184.22	190.34
	5	<i>Empirical</i>	38.1	50.65	61.1	78.88	98.37	119.57	135.67	152.17	169.36
		<i>SW</i>	38.09	50.8	61.27	78.88	98.66	120.13	135.58	151.47	168.11
		<i>MSSD</i>	30.39	41.42	51.21	67.87	87.87	110.19	125.63	144.74	163.54
		<i>Shrinkage</i>	69.86	84.56	93.8	111.72	129.56	146.13	157.71	167.94	181.39

chart. These are evident in the higher in-control ARL for the MCI chart in Tables 2.2-2.4. Also, the ARL_0 values provided in Table 2.5, for $ARL_0 = 500$, confirmed the comparative superiority of the shrinkage estimate over the other estimation methods. In Table 2.5, the benchmark of 500 for the in-control ARL would have been obtained if the known parameters were used to establish control limits.

2.6 Corrected limits of the MCUSUM and MCI charts

As shown in Section 2.5, the control charts from the estimated parameters would require a very large Phase I sample to give performance that rivals the performance of the charts based on known parameters. Specifically, the number of Phase I samples m would need to be much larger than p ,

Table 2.4: In-control ARL of the MCUSUM and MCI control charts for $p = 2$ or 5 , when different Phase I samples (m), of size $n = 1$, are used in estimating the unknown parameters. We used $k = 1$, and the upper control limits that give in-control ARL of 200 when the parameters are known.

<i>Chart</i>	p	<i>Method</i>	m								
			30	40	50	70	100	150	200	300	500
<i>MCUSUM</i>	2	<i>Empirical</i>	134.74	144.01	150.22	158.47	167.15	175.11	180	184.29	188.49
		<i>SW</i>	135.82	145.76	150.35	157.41	167.59	175.53	178.59	184.37	188.28
		<i>MSSD</i>	133.33	144.28	149.72	156.09	166.47	172.95	178.44	184	189.09
		<i>Shrinkage</i>	185.2	185.79	186.3	186.94	187.94	189.58	191.72	192.11	194.29
	5	<i>Empirical</i>	36.27	49.26	60.85	79.64	101.13	123.74	139.94	156.02	172.61
		<i>SW</i>	36.65	49.53	61.85	80.48	101.32	125.22	137.95	155.23	174.4
		<i>MSSD</i>	26.92	37.94	47.52	66.14	86.11	111.8	126.02	148.25	165.76
		<i>Shrinkage</i>	100.44	115.09	123.43	136.16	150.96	164.06	172.57	181.71	188.87
<i>MCI</i>	2	<i>Empirical</i>	132.73	142.03	148.56	160.58	169.61	177.84	183.47	189.94	194.91
		<i>SW</i>	132.09	141.75	148.05	159.73	170.27	178.27	183.97	188.76	195.42
		<i>MSSD</i>	131.13	140.13	145.58	157.00	167.31	174.32	181.28	187.6	193.01
		<i>Shrinkage</i>	188.59	189.8	190.05	191.14	191.69	194.98	196.73	197.97	200.61
	5	<i>Empirical</i>	39.78	54.72	67.55	87.27	108.45	131.52	146.88	163.16	176.03
		<i>SW</i>	39.94	54.96	67.1	87.32	107.84	132.09	146.12	161.95	177.31
		<i>MSSD</i>	27.78	40.35	50.92	69.55	92.62	116.25	131.69	150.82	170.45
		<i>Shrinkage</i>	120.62	133.16	142.97	155.21	168.59	179.11	183.65	191.35	197.9

Table 2.5: In-control ARL of the MCUSUM and MCI control charts for $p = 5$, and some selected values of (m). Here, we used $k = 0.5$, and the upper control limits that give an in-control ARL of 500 when the parameters are known.

<i>Chart</i>	ucl	p	<i>Method</i>	m				
				30	50	100	300	500
<i>MCUSUM</i>	$h_1 = 10.9$	5	<i>Empirical</i>	55.91	92.20	165.71	306.96	364.50
			<i>SW</i>	55.89	92.23	165.75	306.95	364.50
			<i>MSSD</i>	45.42	78.80	148.84	291.69	353.38
			<i>Shrinkage</i>	97.91	143.12	220.92	347.59	393.47
<i>MCI</i>	$h_2 = 7.99$	5	<i>Empirical</i>	60.52	105.64	194.93	340.54	395.09
			<i>SW</i>	60.42	105.45	194.23	338.91	395.55
			<i>MSSD</i>	46.38	86.46	167.96	317.81	374.29
			<i>Shrinkage</i>	126.83	183.9	272.9	392.83	432.35

the number of quality characteristics monitored. Different researchers have recommended different number of Phase I samples when parameters are estimated (e.g., Jones et al., 2001; Champ et al., 2005). However, the recommendation of a very large m relative to p may require a very long

initial monitoring period, during which undetected shifts in the process may occur (Mahmoud and Maravelakis, 2013).

A preferred approach to reduce the frequency of false out-of-control signals of control charts for which the population parameters are unknown is to use wider “corrected” control limits (Aly et al., 2015). This section provides calculated corrected upper control limits: h_{c1} for the MCUSUM chart and h_{c2} for the MCI chart, for the four methods of estimating Σ_0 , using a binary search algorithm (see Champ et al., 2005), along with both Phase I and Phase II analyses explained in Section 2.5. This helps to overcome the variability of the in-control ARL values that often occur due to the use of estimated parameters.

The corrected limits that adjust the in-control ARL values to 200 are given in Tables 2.6 - 2.8. These tables show that the values of h_{c1} and h_{c2} , depend on m , k and the number of quality characteristics being monitored. Smaller values of m give bigger values of h_{c1} and h_{c2} , and as m becomes larger, the values of h_{c1} and h_{c2} approach the corresponding values of h_1 and h_2 , i.e., the limits to be used when the parameters are known (given in Table 2.1, and under the column heading of “ ∞ ” in Tables 2.6 - 2.8). In all of the cases considered, the values of h_{c1} and h_{c2} for charts obtained using the empirical estimate and SW do not differ greatly. Our simulation results show that the values h_{c1} and h_{c2} from the charts based on the shrinkage method are much smaller than those obtained for the other estimation methods considered in this study.

Table 2.6: The corrected upper control limits: (h_{c1}) and (h_{c2}) , that produced an in-control ARL of 200 when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters. The value of k used here is $k = 0.25$.

<i>Chart</i>	<i>p</i>	<i>method</i>	<i>m</i>									
			30	40	50	70	100	150	200	300	500	∞
MCUSUM	2	<i>Empirical</i>	11.76	11.23	10.85	10.32	9.94	9.59	9.37	9.17	8.96	8.64
		<i>SW</i>	11.74	11.22	10.86	10.34	9.93	9.55	9.36	9.16	8.96	
		<i>MSSD</i>	11.79	11.25	10.89	10.39	9.97	9.61	9.38	9.19	8.97	
		<i>Shrinkage</i>	11.17	10.79	10.53	10.12	9.78	9.49	9.30	9.12	8.94	
	3	<i>Empirical</i>	17.69	16.22	15.27	14.2	13.3	12.57	12.18	11.78	11.43	10.85
		<i>SW</i>	17.69	16.27	15.28	14.22	13.32	12.59	12.21	11.77	11.42	
		<i>MSSD</i>	17.98	16.42	15.66	14.42	13.45	12.66	12.24	11.82	11.47	
		<i>Shrinkage</i>	16.02	15.09	14.48	13.67	12.96	12.36	12.03	11.67	11.38	
	5	<i>Empirical</i>	31.24	28.76	25.9	22.73	20.38	18.55	17.65	16.72	15.97	14.81
		<i>SW</i>	31.23	28.78	25.8	22.72	20.37	18.57	17.63	16.73	15.96	
		<i>MSSD</i>	36.3	30.45	27.14	23.5	20.89	18.84	17.84	16.84	15.98	
		<i>Shrinkage</i>	27.72	25.26	23.4	21.12	19.43	17.99	17.23	16.48	15.85	
	2	<i>Empirical</i>	10.11	9.72	9.38	8.94	8.61	8.28	8.13	7.95	7.79	7.49
		<i>SW</i>	10.11	9.71	9.35	8.96	8.6	8.29	8.12	7.94	7.78	
		<i>MSSD</i>	10.2	9.75	9.44	8.97	8.62	8.3	8.15	7.95	7.79	
		<i>Shrinkage</i>	9.66	9.33	9.06	8.72	8.41	8.17	8.03	7.87	7.73	
	3	<i>Empirical</i>	14.07	12.88	12.16	11.2	10.55	9.98	9.67	9.38	9.15	8.75
		<i>SW</i>	14.07	12.83	12.14	11.21	10.51	9.98	9.69	9.38	9.14	
		<i>MSSD</i>	14.62	13.24	12.43	11.45	10.69	10.07	9.73	9.41	9.14	
		<i>Shrinkage</i>	12.68	11.89	11.4	10.78	10.21	9.79	9.52	9.31	9.09	
	5	<i>Empirical</i>	24.38	20.54	18.42	16.04	14.43	13.14	12.55	11.97	11.6	10.80
		<i>SW</i>	24.29	20.61	18.41	16.04	14.39	13.14	12.57	11.97	11.5	
		<i>MSSD</i>	26.75	22.15	19.58	16.91	15.00	13.50	12.87	12.22	11.74	
		<i>Shrinkage</i>	19.55	17.57	16.18	14.65	13.47	12.65	12.17	11.62	11.24	

2.7 Out-of-control ARL performance using corrected limits

In this section, we used simulations to study the zero-state and steady-state out-of-control performance of the control charts, based on the four methods of estimating Σ_0 from Phase I data, for giving an out-of-control signal in Phase II. For the simulations, we followed the same procedures outlined in Section 2.5, including the Phase I sample size and the number of simulations that were run. However, here, the vector of process mean is shifted for Phase II, giving a new vector of

Table 2.7: The corrected upper control limits: (h_{c1}) and (h_{c2}) , that produced an in-control ARL of 200 when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters. The value of k used is $k = 0.5$.

<i>Chart</i>	<i>p</i>	<i>method</i>	m									
			30	40	50	70	100	150	200	300	500	∞
MCUSUM	2	<i>Empirical</i>	6.46	6.31	6.19	6.02	5.9	5.78	5.71	5.63	5.57	5.49
		<i>SW</i>	6.45	6.31	6.19	6.03	5.89	5.77	5.7	5.64	5.58	
		<i>MSSD</i>	6.52	6.38	6.28	6.07	5.9	5.78	5.71	5.65	5.58	
		<i>Shrinkage</i>	6.13	6.05	6.00	5.90	5.81	5.72	5.68	5.62	5.56	
	3	<i>Empirical</i>	9.28	8.78	8.45	8.05	7.72	7.46	7.32	7.17	7.07	6.90
		<i>SW</i>	9.28	8.76	8.44	8.04	7.74	7.46	7.33	7.18	7.06	
		<i>MSSD</i>	9.54	9	8.6	8.18	7.83	7.54	7.36	7.2	7.09	
		<i>Shrinkage</i>	8.26	8.04	7.87	7.68	7.49	7.31	7.22	7.11	7.03	
	5	<i>Empirical</i>	15.97	14.28	13.29	12.18	11.36	10.7	10.38	10.06	9.79	9.40
		<i>SW</i>	15.98	14.29	13.3	12.21	11.36	10.7	10.39	10.05	9.79	
		<i>MSSD</i>	17.42	15.27	14.04	12.67	11.67	10.9	10.52	10.14	9.84	
		<i>Shrinkage</i>	12.87	12.17	11.76	11.16	10.73	10.33	10.1	9.89	9.76	
	2	<i>Empirical</i>	5.68	5.53	5.41	5.28	5.14	5.04	4.99	4.93	4.86	4.78
		<i>SW</i>	5.68	5.52	5.41	5.27	5.15	5.04	4.98	4.92	4.86	
		<i>MSSD</i>	5.69	5.54	5.43	5.29	5.17	5.05	4.99	4.93	4.87	
		<i>Shrinkage</i>	5.31	5.23	5.18	5.08	5.01	4.95	4.9	4.85	4.83	
	3	<i>Empirical</i>	7.45	7.01	6.74	6.41	6.16	5.96	5.84	5.74	5.65	5.52
		<i>SW</i>	7.44	7.02	6.72	6.42	6.17	5.95	5.85	5.73	5.66	
		<i>MSSD</i>	7.74	7.25	6.93	6.55	6.26	6.03	5.9	5.77	5.67	
		<i>Shrinkage</i>	6.55	6.39	6.25	6.08	5.94	5.82	5.76	5.70	5.64	
	5	<i>Empirical</i>	11.42	10.14	9.43	8.62	8.06	7.63	7.41	7.20	7.03	6.81
		<i>SW</i>	11.41	10.15	9.44	8.61	8.06	7.62	7.41	7.19	7.03	
		<i>MSSD</i>	12.6	10.94	10.02	9.04	8.33	7.80	7.54	7.28	7.09	
		<i>Shrinkage</i>	8.92	8.45	8.18	7.81	7.53	7.30	7.18	7.06	6.98	

means: μ_1 . We used shifts of size $\delta = \{0.5, 1, 1.5, 2, 2.5 \text{ or } 3\}$.

The performance of the control charts are evaluated by comparing the out-of-control ARL (denoted by ARL_1) for the different methods. The zero-state ARL_1 is the out-of-control ARL obtained under the assumption that the process shift occurred during the initial stage, while the steady-state ARL_1 is the out-of-control ARL under the assumption that the process shift occurred after the process had been in control for some time (Lucas and Saccucci, 1990; Golosnoy et al., 2009; Kim, 2014; Wang and Huang, 2016). For the steady-state, we adopted the scheme employed by Siegmund (1985), and Ngai and Zhang (2001), by assuming that for a fixed integer q , the

Table 2.8: The corrected upper control limits: (h_{c1}) and (h_{c2}) , that produced an in-control ARL of 200 when different Phase I samples (m) , of size $n = 1$ are used in estimating the unknown parameters. The value of k used is $k = 1$.

<i>Chart</i>	<i>p</i>	<i>method</i>	m									
			30	40	50	70	100	150	200	300	500	∞
MCUSUM	2	<i>Empirical</i>	3.22	3.19	3.18	3.14	3.13	3.1	3.09	3.06	3.05	3.00
		<i>SW</i>	3.21	3.19	3.18	3.14	3.13	3.09	3.08	3.07	3.05	
		<i>MSSD</i>	3.28	3.22	3.18	3.15	3.12	3.1	3.08	3.07	3.05	
		<i>Shrinkage</i>	3.07	3.06	3.05	3.05	3.05	3.05	3.04	3.04	3.03	
	3	<i>Empirical</i>	4.54	4.37	4.27	4.13	4.03	3.95	3.91	3.86	3.82	3.77
		<i>SW</i>	4.52	4.37	4.27	4.12	4.03	3.95	3.9	3.85	3.82	
		<i>MSSD</i>	4.67	4.47	4.35	4.21	4.09	3.98	3.93	3.88	3.83	
		<i>Shrinkage</i>	3.96	3.95	3.94	3.91	3.88	3.86	3.85	3.83	3.81	
	5	<i>Empirical</i>	7.46	6.88	6.54	6.16	5.86	5.63	5.52	5.41	5.32	5.21
		<i>SW</i>	7.47	6.89	6.55	6.15	5.85	5.63	5.52	5.4	5.31	
		<i>MSSD</i>	8.16	7.39	6.94	6.43	6.04	5.76	5.62	5.47	5.36	
		<i>Shrinkage</i>	5.84	5.73	5.65	5.54	5.46	5.38	5.35	5.3	5.27	
MCI	2	<i>Empirical</i>	2.96	2.92	2.89	2.86	2.82	2.79	2.78	2.75	2.74	2.71
		<i>SW</i>	2.96	2.92	2.89	2.85	2.82	2.79	2.78	2.75	2.74	
		<i>MSSD</i>	2.99	2.94	2.9	2.86	2.83	2.8	2.77	2.76	2.74	
		<i>Shrinkage</i>	2.77	2.75	2.74	2.73	2.72	2.72	2.71	2.7	2.69	
	3	<i>Empirical</i>	3.82	3.67	3.57	3.46	3.37	3.29	3.26	3.22	3.19	3.14
		<i>SW</i>	3.83	3.67	3.58	3.46	3.37	3.29	3.25	3.21	3.19	
		<i>MSSD</i>	3.99	3.83	3.68	3.53	3.42	3.33	3.28	3.23	3.21	
		<i>Shrinkage</i>	3.31	3.29	3.27	3.25	3.23	3.21	3.20	3.19	3.18	
	5	<i>Empirical</i>	5.66	5.19	4.91	4.6	4.37	4.22	4.13	4.06	4.01	3.92
		<i>SW</i>	5.66	5.18	4.9	4.62	4.38	4.21	4.14	4.06	4.01	
		<i>MSSD</i>	6.31	5.65	5.27	4.85	4.55	4.32	4.21	4.10	4.05	
		<i>Shrinkage</i>	4.3	4.23	4.17	4.11	4.06	4.01	3.99	3.97	3.95	

process is under control for the first $q - 1$ samples, and out-of-control after the $(q - 1)th$ sample.

Hence, the steady-state out-of-control ARL is obtained by subtracting $q - 1$ from a simulated run length (Ngai and Zhang, 2001). Without loss of generality, we used $q = 16$. Table 2.9 and Table 2.10 present the zero-state and steady-state ARL_1 values, respectively, of both the MCUSUM and MCI charts when $p = 5$, $m = \{30, 50 \text{ or } 200\}$, and $k = 0.5$. In both tables, the corrected control limits of the competing charts are chosen to give an ARL_0 of 200 when there was no shift in the process. The ARL_1 values of both charts for the known parameter case are given in the rows with $m = \infty$.

Table 2.9: The zero-state ARL_1 of the MCUSUM and MCI charts from different methods of estimating the covariance matrix, for $p = 5$, different values of m and $k = 0.5$.

		δ						
	Methods	m	0.5	1	1.5	2	2.5	3
<i>MCUSUM</i>	<i>Empirical</i>	30	63.28	21.45	12.91	9.32	7.33	6.04
	<i>SW</i>		63.21	21.45	12.96	9.34	7.3	6.03
	<i>MSSD</i>		64.22	22.07	13.38	9.66	7.59	6.28
	<i>Shrinkage</i>		57.66	18.96	11.39	8.18	6.43	5.32
	<i>Empirical</i>	50	52.73	18.56	11.25	8.11	6.38	5.28
	<i>SW</i>		53.18	18.57	11.28	8.12	6.39	5.29
	<i>MSSD</i>		53.68	19.05	11.54	8.34	6.56	5.43
	<i>Shrinkage</i>		49.88	17.32	10.48	7.53	5.92	4.92
	<i>Empirical</i>	200	39.85	14.92	9.16	6.65	5.25	4.37
	<i>SW</i>		39.81	14.97	9.18	6.65	5.25	4.37
	<i>MSSD</i>		40.1	15.02	9.21	6.7	5.29	4.39
	<i>Shrinkage</i>		39.62	14.75	9.02	6.55	5.17	4.3
	<i>known parameter</i>	∞	35.57	13.57	8.38	6.14	4.84	4.04
<i>MCI</i>	<i>Empirical</i>	30	56.43	17.15	9.7	6.85	5.37	4.45
	<i>SW</i>		55.08	17.23	9.65	6.86	5.38	4.45
	<i>MSSD</i>		58.19	17.92	10.1	7.2	5.65	4.66
	<i>Shrinkage</i>		49.48	14.75	8.25	5.86	4.61	3.82
	<i>Empirical</i>	50	45.86	14.78	8.36	5.95	4.67	3.88
	<i>SW</i>		46.01	14.78	8.37	5.95	4.68	3.89
	<i>MSSD</i>		47.2	15.17	8.59	6.13	4.81	4.00
	<i>Shrinkage</i>		42.65	13.47	7.62	5.43	4.27	3.56
	<i>Empirical</i>	200	38.83	11.9	6.87	4.92	3.89	3.24
	<i>SW</i>		38.67	11.95	6.86	4.91	3.88	3.24
	<i>MSSD</i>		38.06	12.08	6.95	4.96	3.92	3.28
	<i>Shrinkage</i>		37.69	11.69	6.73	4.83	3.82	3.18
	<i>known parameter</i>	∞	35.32	11.01	6.39	4.59	3.64	3.03

For all of the estimation methods considered here, the results in Tables 2.9 and 2.10 show increases in the ARL_1 for charts relying on estimated parameters compared to charts with known parameters. However, the results show that the shrinkage estimate has an even better out-of-control ARL performance than any of the other methods considered. As expected, the MCI chart gives better zero-state ARL performance than the MCUSUM chart. This is evident in the lower ARL_1 values for the MCI chart in Table 2.9. However, the MCI chart shows poor steady-state ARL performance (see Table 2.10). We refer interested readers to Crosier (1988), Pignatiello and Runger (1990), Ngai and Zhang (2001), Golosnoy et al. (2009), Kim (2014), and Wang and

Table 2.10: The steady-state ARL_1 of the MCUSUM and MCI charts from different methods of estimating the covariance matrix, for $p = 5$, different values of m and $k = 0.5$.

		δ						
	Methods	m	0.5	1	1.5	2	2.5	3
MCUSUM	<i>Empirical</i>	30	54.75	18.62	11.37	8.18	6.47	5.39
	<i>SW</i>		57.63	18.69	11.28	8.22	6.47	5.4
	<i>MSSD</i>		56.51	19.11	11.69	8.51	6.72	5.62
	<i>Shrinkage</i>		50.00	16.31	9.86	7.13	5.63	4.68
	<i>Empirical</i>	50	47.59	16.26	10	7.23	5.71	4.74
	<i>SW</i>		47.08	16.43	9.9	7.19	5.72	4.76
	<i>MSSD</i>		47.5	16.68	10.23	7.39	5.86	4.87
	<i>Shrinkage</i>		43.99	15.25	9.21	6.68	5.28	4.38
	<i>Empirical</i>	200	36.04	13.25	8.16	5.93	4.68	3.92
	<i>SW</i>		35.9	13.3	8.2	5.94	4.7	3.91
	<i>MSSD</i>		36.54	13.32	8.19	5.96	4.72	3.94
	<i>Shrinkage</i>		35.6	13.12	8.04	5.85	4.61	3.85
	known parameter	∞	31.58	12.09	7.4	5.42	4.37	3.4
MCI	<i>Empirical</i>	30	61.41	18.78	11.3	8.3	6.78	5.77
	<i>SW</i>		63.2	18.59	11.38	8.37	6.77	5.81
	<i>MSSD</i>		63.07	19.15	11.51	8.55	6.95	5.87
	<i>Shrinkage</i>		56.19	16.22	9.83	7.34	5.96	5.05
	<i>Empirical</i>	50	53.19	16.82	10.22	7.59	6.16	5.26
	<i>SW</i>		53.37	16.61	10.18	7.59	6.16	5.27
	<i>MSSD</i>		54.16	17.87	10.38	7.7	6.24	5.28
	<i>Shrinkage</i>		50.78	15.6	9.44	7.04	5.74	4.88
	<i>Empirical</i>	200	42.5	14.09	8.74	6.54	5.33	4.56
	<i>SW</i>		42.21	14.16	8.75	6.53	5.3	4.56
	<i>MSSD</i>		43.01	14.31	8.81	6.61	5.35	4.57
	<i>Shrinkage</i>		41.48	14	8.56	6.42	5.25	4.51
	known parameter	∞	38.86	13.39	8.14	6.17	5.11	4.27

Huang (2016), for the zero-state and steady-state properties and comparisons of MCUSUM and MCI control charts.

2.8 Estimation of the corrected limits from the shrinkage estimate

We estimated the corrected limits from the charts based on the shrinkage estimate of Σ_0 , using linear models so that control-chart practitioners may obtain estimated corrected limits when using

the shrinkage method beyond the values of m considered in this study. The least squares models are given in Tables 2.11 and 2.12 for the MCUSUM and MCI control charts, where the logarithm of m (with base 10) are used as the explanatory variables in the models. These models are obtained from the values of the corrected upper control limits, h_{c1} and h_{c2} , given in Tables 2.6 - 2.8. By using these estimates, a chart with $ARL_0 \approx 200$ should be obtained. The coefficients of determination R^2 were at least 86% and 89% for the linear models for MCUSUM and MCI control charts, respectively.

Table 2.11: Simple linear models that may be used to estimate the corrected upper control limit h_{c1} that will give an in-control $ARL \approx 200$ for a shrinkage-MCUSUM control chart, for different values of p and reference parameter k . The coefficient of determination for each model is given in parentheses.

p	k					
	0.25		0.5		1	
2	$13.67 - 1.85 \times \log(m)$	(95%)	$6.80 - 0.48 \times \log(m)$	(98%)	$3.10 - 0.03 \times \log(m)$	(86%)
3	$20.99 - 3.79 \times \log(m)$	(94%)	$9.62 - 1.02 \times \log(m)$	(95%)	$4.15 - 0.13 \times \log(m)$	(98%)
5	$40.03 - 9.62 \times \log(m)$	(91%)	$16.10 - 2.53 \times \log(m)$	(92%)	$6.45 - 0.47 \times \log(m)$	(93%)

Table 2.12: Simple linear models that may be used to estimate the corrected upper control limit h_{c2} that will give an in-control $ARL \approx 200$ for a shrinkage-MCI control chart, for different values of p and reference parameter k . The coefficient of determination for each model is given in parentheses.

p	k					
	0.25		0.5		1	
2	$5.86 - 0.41 \times \log(m)$	(95%)	$5.86 - 0.41 \times \log(m)$	(95%)	$2.84 - 0.10 \times \log(m)$	(90%)
3	$16.41 - 2.90 \times \log(m)$	(92%)	$7.54 - 0.76 \times \log(m)$	(93%)	$3.45 - 0.11 \times \log(m)$	(95%)
5	$27.59 - 6.56 \times \log(m)$	(89%)	$10.87 - 1.55 \times \log(m)$	(90%)	$4.67 - 0.28 \times \log(m)$	(91%)

2.9 Illustrative example

We illustrate the performance of the shrinkage method using the bimetal thermostat dataset taken from Santos-Fernandez (2012). Bimetal thermostats have a bimetallic strip, composed of two strips

of different metals, which convert changes in temperature to a mechanical displacement *via* thermal expansion, so that the temperature may be read (Santos-Fernandez, 2012). The dataset contains measurements of the deflection, curvature, resistivity, and hardness: in low and high expansion sides, of brass and steel bimetal thermostats. Data for both Phase I and Phase II are used, each of size $m = 28$, and with $p = 5$ variables. The process parameters $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$ were estimated from the Phase I data. We applied each of the four methods for estimating $\boldsymbol{\Sigma}_0$ and compared their performances. We used $k = 0.5$ with two different control limits: the upper control limits h that give an in-control ARL of 200 when parameters are known (Table 2.1), and the corrected upper control limits h_{c1} and h_{c2} that give an in-control ARL of 200 (Table 2.7) when parameters are unknown.

Results were very consistent across the methods when using the upper control limit h ; for both the MCUSUM and MCI control charts: the empirical estimate, SW, and shrinkage estimate methods all detected the signal after the 11th observation, whereas the MSSD method detected it after the 12th observation (Figures 2.1 and 2.2). In contrast, for the MCUSUM control charts with the corrected upper control limits h_{c1} , the shrinkage estimate method detected the signal after the 19th observation, the empirical estimate and the SW methods both detected the signal after the 23rd observation, and the MSSD method failed to detect the signal at all. Similarly, for the MCI control charts with h_{c2} limits, the shrinkage estimate method detected the first signal after the 14th observation, the empirical estimate and SW methods both detected the signal after the 23rd observation, and the MSSD method failed to detect the signal. We recommend the use of h_{c1} , and h_{c2} rather than h for this example, because the Phase II development of the charts used estimated rather than known parameters.

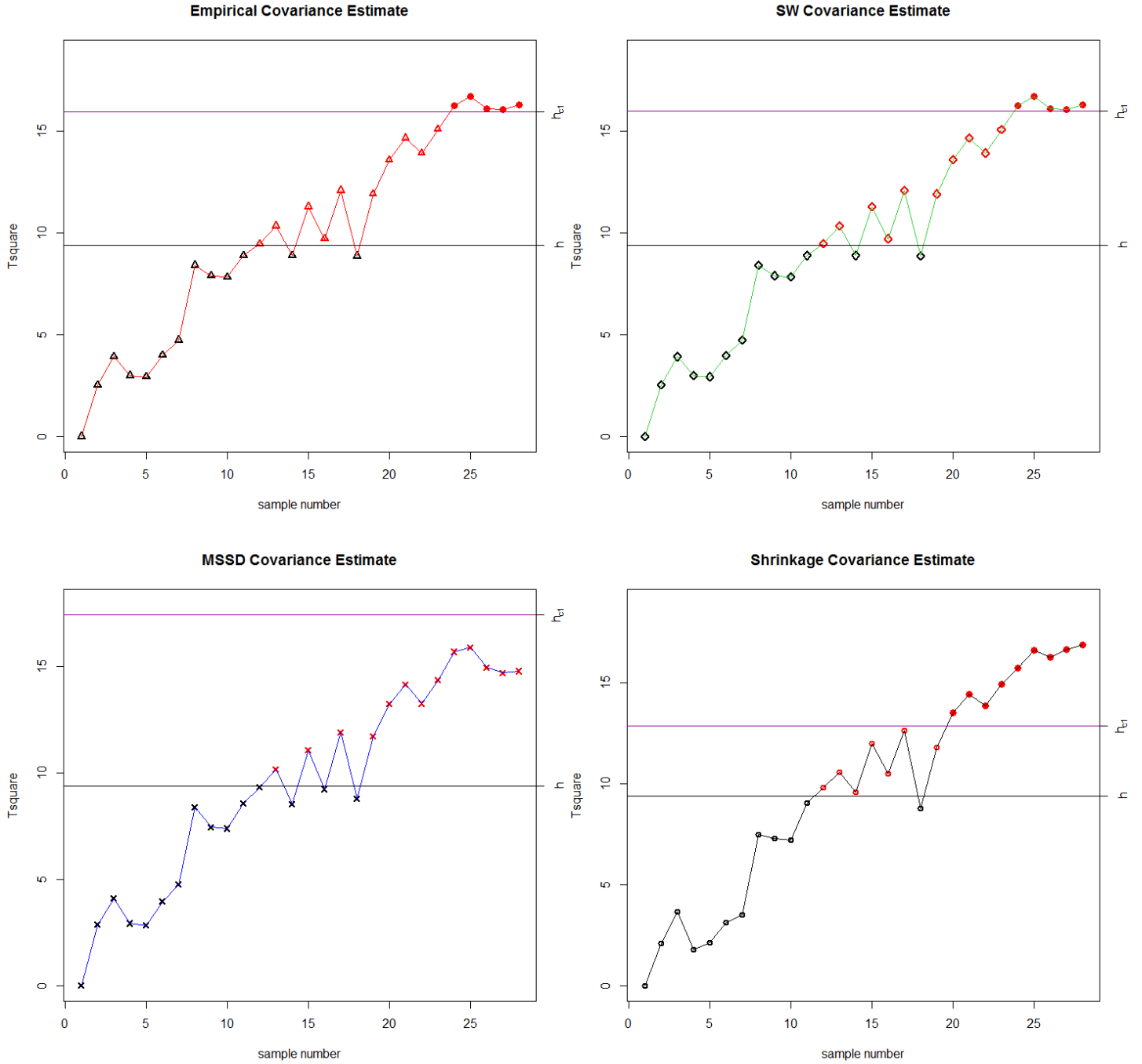


Figure 2.1: The MCUSUM control chart of the Bimetal II data set. The estimated covariance matrices are obtained from the Bimetal I data set. The reference parameter used is $k = 0.5$ and $p = 5$. The upper control limit h , for the known parameters case (Table 2.1) and the corrected upper control limit h_{c1} ; (Table 2.7) used here are the ones that give an in-control ARL of 200.

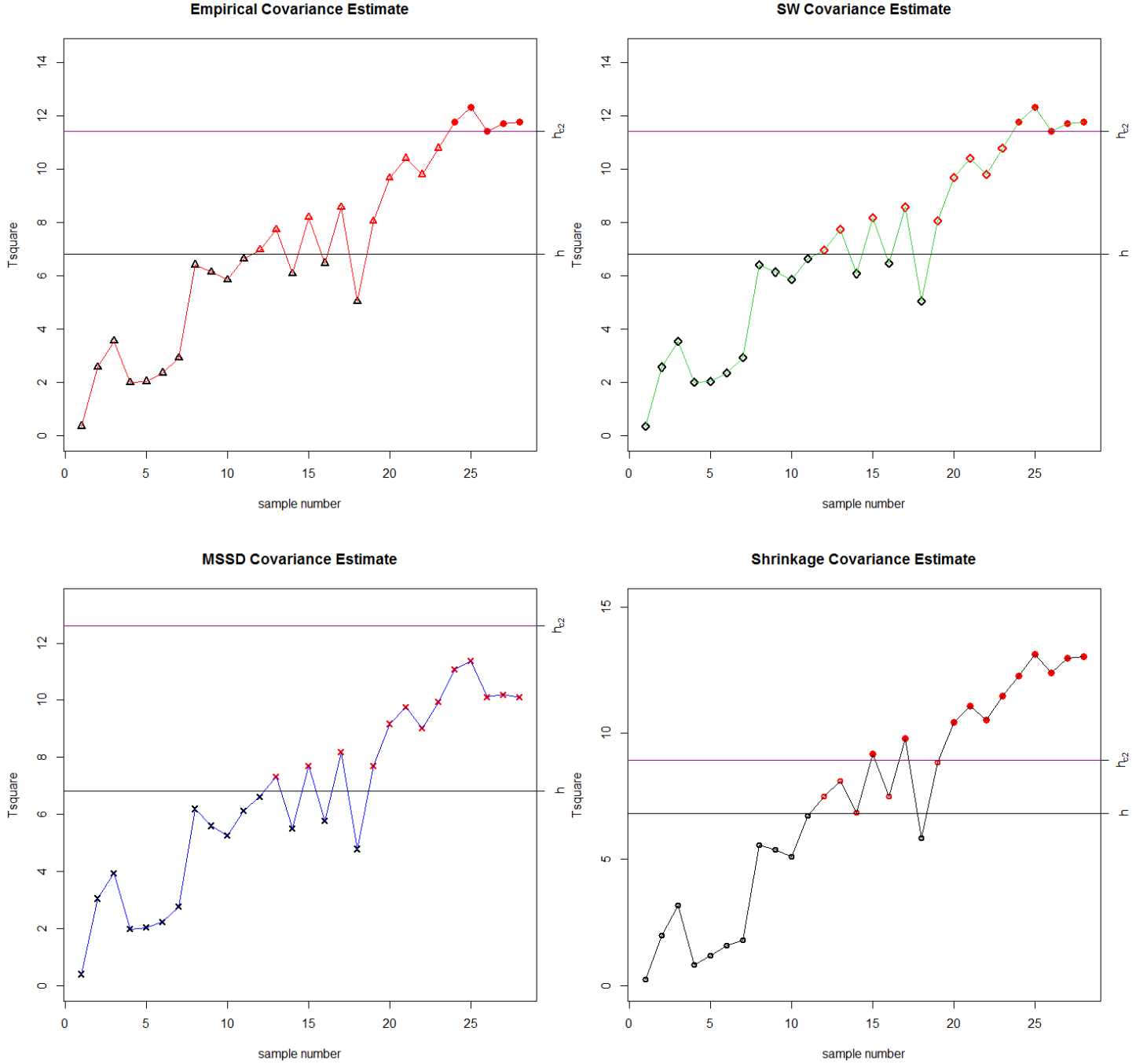


Figure 2.2: The MCI control chart of the Bimetal II data set. The estimated covariance matrices are obtained from the Bimetal I data set. The reference parameter used is $k = 0.5$ and $p = 5$. The upper control limit h , for the known parameters case (Table 2.1) and the corrected upper control limit h_{c1} (Table 2.7) used here are the ones that give an in-control ARL of 200.

2.10 General conclusions

In this chapter, we investigated the performance of the multivariate cumulative sum control chart

methods, MCUSUM and MCI control charts, for individual-observation monitoring when process

parameters are unknown and estimated from Phase I data. We propose the use of a shrinkage estimate for the covariance matrix and present the results of simulation studies to compare the performance of control charts obtained using this estimate versus those obtained using several alternative methods for estimating the covariance matrix.

We demonstrate that, if the upper control limits of the charts are used, the in-control performance of the MCUSUM and MCI control charts are negatively affected by estimating the process parameters unless a large Phase I sample is available. For different Phase I sample sizes (m), we calculated the values of the corrected control limits that give the desired in-control ARL when using estimated parameters, and have provided least-squares models that can be used to estimate the corrected limits for values of m that were not considered in this study. Our simulations showed that the control charts obtained using the shrinkage estimate of the covariance matrix are superior to those obtained using other methods of estimating the covariance matrix. The relative performance of the control charting schemes, including the superiority of the shrinkage estimation method, was also readily illustrated in a particular example - the analysis of a multivariate Bimetal dataset.

Chapter 3

Multivariate exponentially weighted moving average (MEWMA) control charts for individual-observations monitoring when parameters are estimated

This chapter has been submitted for publication

Multivariate exponentially weighted moving average (MEWMA) control charts for individual monitoring require *a priori* knowledge of the in-control parameters. In practice, this assumption is not always tenable, and estimated parameters are generally obtained from an in-control reference sam-

ple of m preliminary observations (Phase I sample). Here, we compared the Phase II performance of MEWMA control charts using different methods for estimating the covariance matrix when the in-control covariance structure is unknown, and only a small Phase I sample (m) is available. The performance of the MEWMA control charts varied among the methods. For simulated data with small m , the performance of MEWMA control charts using a shrinkage estimate of the covariance matrix was superior to the alternative methods considered in this study. Specifically, the MEWMA chart obtained using the shrinkage estimate gave a longer average run length (ARL) for in-control processes and shorter ARL for out-of-control processes. The improved performance of MEWMA control charts using the shrinkage estimate was also demonstrated *via* illustrative case studies of bimetal thermostat and gene expression applications; changes were detected earlier by the shrinkage-based MEWMA method.

3.1 Introduction

Control charts are statistical process-control (SPC) methods, used to monitor and detect changes in the uniformity of a process (Aslam et al., 2015; Khan et al., 2017). Control charts typically involve two phases (Jensen et al., 2006). In Phase I, a reference sample of measurements from a process known to be in control is analyzed to establish the expected behavior of the process, estimate the process population parameters (if they are unknown), and determine control limits for a reference statistic. Subsequent measurements are then collected during Phase II and monitored for significant deviations from the in-control state. Any values of the statistic outside the control limits established in Phase I generate an out-of-control signal, prompting some remedial response (Abbasi and Adegoke, 2018). A common measure used to evaluate the performance of control

charts is the “average run length” (ARL), given as the average number of Phase II iterations that occur before the first out-of-control signal occurs (Riaz et al., 2014; Khaliq et al., 2016). An efficient control chart maximizes the ARL when the process is in control while minimizing the ARL when the process is out of control (Jones et al., 2001).

Control chart methods can be classified broadly into two types (Sanusi et al., 2017b; Ajadi et al., 2016). Firstly, memoryless control charts, such as the Shewhart method Shewhart (1931), evaluate data at only a single time point, disregarding any past behavior of the process during Phase II. These methods are typically used where any shifts in the process mean are expected to be relatively large. In contrast, memory-based control charts use both the current and recent data to provide greater power to identify small to moderate shifts. Examples include the exponentially weighted moving average (EWMA) (Roberts, 1959), the cumulative sum (CUSUM) (Page, 1954), the homogeneously weighted moving average (HWMA) (Abbas, 2018), and their multivariate counterparts, i.e., the MEWMA chart (Lowry et al., 1992), the MCUSUM charts (Crosier, 1988; Pignatiello and Runger, 1990), and the MHWMA chart Adegoke et al. (2019), respectively. Because the memory-type control charts combine information from several time points, they are particularly effective at detecting small to moderate shifts in the process, and are highly efficient for individual-observation ($n = 1$) processes (Montgomery, 2009).

Multivariate control chart analyses are relatively simple to implement in cases where the in-control process parameters are known *a priori*, allowing desired control limits to be calculated with certainty. However, in most cases, the process parameters are unknown and control limits are calculated based on parameters that have been estimated from Phase I data. The uncertainty introduced by estimating the process parameters typically reduces the in-control ARL of the charts.

Several authors have investigated the performance of the multivariate Shewhart control chart for individual-observation monitoring, when parameters are unknown and are estimated from Phase I. An empirical estimate of the in-control covariance matrix (Σ_0) may be calculated directly from the m multivariate Phase I data points (Montgomery, 2009). Sullivan and Woodall (1996) showed that the empirical estimate of Σ_0 leads to poor performance of the Hotelling T^2 control chart for detecting shifts in the mean vector. Holmes and Mergen (1993) showed that the empirical estimate of Σ_0 is insensitive to out-of-control shifts, and proposed the use of a mean square successive difference (MSSD) approach to estimate the variance-covariance matrix. In all of these studies, the effects of having to estimate the in-control parameters may be minimized by having a large Phase I dataset (m) (Jones et al., 2001; Mahmouda and Maravelakisb, 2010).

Different methods for estimating Σ_0 have been used with MEWMA control charts for individual-observation monitoring Zhang and Chang (2008); Aly et al. (2016). Yet, the performance of MEWMA charts for individual-observation monitoring when parameters are estimated have not been directly compared in the SPC literature. Such comparisons are required in order to determine the most appropriate methods for estimating the covariance parameters, and to make recommendations to practitioners, particularly when the number of Phase I samples m is small. In this study, we investigated the performance of the MEWMA control chart for individual-observation monitoring when the process parameters (i.e., mean vector μ_0 and the covariance matrix Σ_0) are to be estimated from a small number of Phase I samples (m).

We compared the performance of MEWMA control charts using three different methods of estimation for Σ_0 . Two methods are currently in use: the empirical estimate and the MSSD. The third method, proposed here, is a shrinkage estimate, given as a weighted combination of the

empirical estimate and a target matrix. This study provides a guide for practitioners to make an appropriate choice regarding the estimation of Σ_0 for MEWMA control charts for individual-observation monitoring when a large Phase I dataset is not available. We refer the reader to Champ and Jones-Farmer (2007); Aly et al. (2016); Saleh, Nesma A and Mahmoud (2017), and references therein, for comparisons of the performance of the MEWMA control charts based on estimated parameters when multiple observations are taken at each time point (i.e., $n > 1$).

The article is organized as follows. In Section 3.2, we describe the MEWMA control chart, review methods of estimating parameters for individual-observation monitoring, and propose a shrinkage estimate of Σ_0 in Section 3.3. Section 3.4 compares the performance of MEWMA control charts when Σ_0 is estimated using different methods, by comparing their in-control ARL. Corrected control limits for the MEWMA chart for individual-observation monitoring when estimation of Σ_0 is required are presented in Section 3.4.2, followed by a comparison of the overall out-of-control ARL performance under these corrected limits, using simulated data, in Section 3.4.2. In Section 3.5, we show examples of the applications of the methods, and a discussion is given in Section 3.6.

3.2 Multivariate exponentially weighted moving average control chart (MEWMA)

3.2.1 General framework

Rapid development in data-acquisition and online monitoring have led to increased interest in multivariate control chart methods, which monitor several features of a process simultaneously (Ou et al., 2015). For most practical applications, a $p \times 1$ vector \mathbf{X} of p quality characteristics

is recorded from a single randomly sampled unit at each of m regularly spaced time points. In some cases, $n > 1$ are available at each time point (Mahmouda and Maravelakisb, 2010). Here, we consider only the individual-observation case, where $n = 1$. The \mathbf{X} are assumed to be independent and identically distributed as multivariate normal (MVN) random variables with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$ (Montgomery, 2009).

The MEWMA control chart, proposed by Lowry et al. (1992), is a multivariate extension of the EWMA chart, a memory-type method that accumulates information across time points. Let the p monitored features at time-point i be represented by a $p \times 1$ random vector, \mathbf{x}_i . The MEWMA statistics are given as:

$$\mathbf{w}_i = \mathbf{R}\mathbf{x}_i + (\mathbf{I} - \mathbf{R})\mathbf{w}_{i-1}, \quad (3.1)$$

where, $i = 1, 2, \dots, t$, $\mathbf{w}_0 = \boldsymbol{\mu}_0$ and \mathbf{R} is a $p \times p$ diagonal square matrix with smoothing constants r_k , $k = 1, 2, \dots, p$ along the diagonal such that $0 < r_k \leq 1$, which determines the relative importance of “older” data towards the calculation of the MEWMA statistic. The matrix \mathbf{I} is a diagonal matrix of 1's on its main diagonal. The chart gives an out-of-control signal when:

$$T_i^2 = (\mathbf{w}_i - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}_i^{-1} (\mathbf{w}_i - \boldsymbol{\mu}_0) > h, \quad (3.2)$$

where h and \mathbf{R} are chosen to achieve a desired in-control performance measure (such as, a desired value of in-control ARL), and $\boldsymbol{\Sigma}_i$ is the covariance matrix at time point i . In practice, there is generally no reason to employ different values of the smoothing parameters at different time points (Lowry et al., 1992). If the values of the smoothing parameter, which determine the weight of each prior observation, are equal across variables, i.e., $r_1 = r_2 = \dots = r_p = r$, then the MEWMA vector

becomes:

$$\mathbf{w}_i = r\mathbf{x}_i + (1 - r)\mathbf{w}_{i-1}, \quad (3.3)$$

Lowry et al. (1992) provided two alternative forms of Σ_i : the exact covariance matrix is given as $\Sigma_i = \frac{r[1 - (1 - r)^{2i}]}{2 - r}\Sigma_0$, and the asymptotic covariance matrix is given as $\Sigma = \frac{r}{2 - r}\Sigma_0$. Using simulated data, they showed that the MEWMA chart is more efficient in detecting small to moderate shifts than a memoryless control chart based on the Hotelling T^2 statistic. The MEWMA chart is a directionally invariant chart. As such, the performance of the chart depends on the mean vector and the covariance matrix only through the non-centrality parameter δ (Alkahtani and Schaffer, 2012), given as:

$$\delta = \sqrt{(\boldsymbol{\mu} - \boldsymbol{\mu}_0)^T \Sigma_0^{-1} (\boldsymbol{\mu} - \boldsymbol{\mu}_0)}. \quad (3.4)$$

where $\boldsymbol{\mu}$, is the location parameter for the out-of-control process. The use of small values for the smoothing parameter increases the power of the control chart. If $r = 1$, the chart is identical to the memoryless control chart based on Hotelling's T^2 .

3.3 Estimating the Phase I parameters

Here, we provide details of the methods used in this study to estimate the mean vector and covariance matrix. We assume that the population parameters, $\boldsymbol{\mu}_0$ and Σ_0 , are to be estimated from a Phase I sample values, x_{ki} , with a single observation ($n = 1$) taken for each of $k = 1, \dots, p$ variables at each of $i = 1, \dots, m$ time points.

3.3.1 Estimating the mean vector

For all methods herein, the empirical estimate of the mean vector $\boldsymbol{\mu}_0$, is obtained by taking the average across m Phase I in-control samples, namely $\bar{\mathbf{x}}$ is a $p \times 1$ length vector of elements:

$$\bar{x}_k = \frac{1}{m} \sum_{i=1}^m x_{ki}. \quad (3.5)$$

3.3.2 Estimating the covariance matrix

Empirical estimate

The classical unbiased empirical estimate of the covariance matrix, $\boldsymbol{\Sigma}_0$, of the p -variate random variables is given as:

$$\mathbf{S} = \frac{1}{m-1} \sum_{i=1}^m (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T, \quad (3.6)$$

where the main diagonal of \mathbf{S} contains the sample variances $s_k^2 = \frac{1}{m-1} \sum_{i=1}^m (x_{ik} - \bar{x}_k)^2$, and the off diagonal elements of \mathbf{S} are the sample covariances given as: $s_{kk'} = \frac{1}{m-1} \sum_{i=1}^m (x_{ik} - \bar{x}_k)(x_{ik'} - \bar{x}_{k'})$, for $k = 1, \dots, p$, $k' = 1, \dots, p$ and $k \neq k'$ (Montgomery, 2009).

Mean Square Successive Difference (MSSD)

To make the Hotelling T^2 control chart more sensitive to process changes, Holmes and Mergen (1993) proposed the mean square successive difference (MSSD) approach for estimating the variance-covariance matrix. This method uses the differences between consecutive observations:

$$\mathbf{v}_i = \mathbf{x}_{i+1} - \mathbf{x}_i, i = 1, 2, \dots, m-1 \quad (3.7)$$

The MSSD estimator of Σ_0 is one-half the sample covariance matrix of the matrix of the differences, and is given as:

$$\hat{\Sigma}_{MSSD} = \frac{\mathbf{V}^T \mathbf{V}}{2(m-1)}, \quad (3.8)$$

where \mathbf{V} is given as:

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_{(m-1)}^T \end{bmatrix}$$

Shrinkage estimate

The shrinkage estimator is a weighted average of the empirical unbiased estimate \mathbf{S} given in equation (3.6), and a target matrix $\mathbf{T} = (t_{ik})_{1 \leq i, k \leq p}$, and is given as:

$$\hat{\Sigma}_S = \lambda \mathbf{T} + (1 - \lambda) \mathbf{S}, \quad (3.9)$$

where, $\lambda \in [0, 1]$ denotes the shrinkage intensity: $\lambda = 0$ implies $\hat{\Sigma} = \mathbf{S}$, and $\lambda = 1$ gives $\hat{\Sigma} = \mathbf{T}$.

The estimator is better conditioned than the empirical estimate of the covariance matrix, and asymptotic properties hold well even with finite samples (Warton, 2008; Ledoit and Wolf, 2004a). It does not make any distributional assumption about the data, such as multivariate normality (Ullah et al., 2017).

Several studies have proposed different methods to estimate the optimal value (i.e., λ^*) of the shrinkage intensity parameter (λ) given in Equation (3.9). For example, Ledoit and Wolf (2004a), and Schäfer and Strimmer (2005) gave analytic approaches for determining λ^* . Friedman

(1989) proposed a computationally intensive approach to estimate λ^* using cross-validation. Morris (1983), and Greenland (2000) estimate λ^* in an empirical Bayes context. Here, we followed the approach in Ledoit and Wolf (2004a), and Schäfer and Strimmer (2005), and derived the optimal intensity parameter analytically. The optimal shrinkage intensity, λ^* , is obtained consistently by considering the Frobenius norm of the difference between the shrinkage estimator and the true covariance matrix given as:

$$L(\lambda) = \|\hat{\Sigma}_S - \Sigma_0\|^2 \quad (3.10)$$

The Frobenius norm difference is a quadratic measure of distance between the true and the estimated covariance matrices, (i.e., a loss function). Using Equation (3.9), Equation (3.10) can be written as:

$$L(\lambda) = \|\lambda \mathbf{T} + (1 - \lambda) \mathbf{S} - \Sigma_0\|^2 \quad (3.11)$$

The squared Frobenius norm of the $p \times p$ symmetric matrix \mathbf{T} with entries $(t_{kk'})$ where k and $k' = 1, \dots, p$ is given as:

$$\|\mathbf{T}\|^2 = \sum_{k=1}^p \sum_{k'=1}^p t_{kk'}^2 \quad (3.12)$$

where $\|\cdot\|^2$ is the squared Frobenius norm, which is a quadratic form associated with the inner product Strang (2009). Hence, the loss function in Equation (3.11) becomes:

$$L(\lambda) = \sum_{k=1}^p \sum_{k'=1}^p (\lambda t_{kk'} + (1 - \lambda) s_{kk'} - \sigma_{kk'})^2$$

where, $t_{kk'}$, $s_{kk'}$, and $\sigma_{kk'}$ are the (k, k') th entry of \mathbf{T} , \mathbf{S} , and Σ_0 , respectively. The optimal shrinkage intensity (λ^*), is obtained by minimizing the expected value of $L(\lambda)$ given as:

$$R(\lambda) = E(L(\lambda)) = \sum_{k=1}^p \sum_{k'=1}^p E(\lambda t_{kk'} + (1 - \lambda)s_{kk'} - \sigma_{kk'})^2 = \sum_{k=1}^p \sum_{k'=1}^p (\text{MSE}(\lambda t_{kk'} + (1 - \lambda)s_{kk'}))$$

$$R(\lambda) = \sum_{k=1}^p \sum_{k'=1}^p (\lambda^2 \text{Var}(t_{kk'}) + (1 - \lambda)^2 \text{Var}(s_{kk'}) + 2\lambda(1 - \lambda) \text{Cov}(t_{kk'}, s_{kk'}) + (\lambda E(t_{kk'} - s_{kk'}) + \text{bias}(s_{kk'}))^2)$$

where $\text{bias}(s_{kk'}) = 0$

$$R(\lambda) = \sum_{k=1}^p \sum_{k'=1}^p (\lambda^2 \text{Var}(t_{kk'}) + (1 - \lambda)^2 \text{Var}(s_{kk'}) + 2\lambda(1 - \lambda) \text{Cov}(t_{kk'}, s_{kk'}) + \lambda^2 (E(t_{kk'} - s_{kk'}))^2)$$

The optimal shrinkage parameter is obtained by minimizing the risk function. The goal is to minimize $R(\lambda)$ with respect to λ . This is done by calculating the first two derivatives of $R(\lambda)$ subject to λ . The first derivative is given as:

$$R'(\lambda) = 2 \sum_{k=1}^p \sum_{k'=1}^p (\lambda \text{Var}(t_{kk'}) - (1 - \lambda) \text{Var}(s_{kk'}) + (1 - 2\lambda) \text{Cov}(t_{kk'}, s_{kk'}) + \lambda (E(t_{kk'} - s_{kk'}))^2)$$

Setting $R'(\lambda) = 0$, we have:

$$\lambda \sum_{k=1}^p \sum_{k'=1}^p \text{Var}(t_{kk'} - s_{kk'}) + (E(t_{kk'} - s_{kk'}))^2 = \sum_{k=1}^p \sum_{k'=1}^p \text{Var}(s_{kk'}) - \text{Cov}(t_{kk'}, s_{kk'}),$$

and solving for λ gives:

$$\lambda^* = \frac{\sum_{k=1}^p \sum_{k'=1}^p \text{Var}(s_{kk'}) - \text{Cov}(t_{kk'}, s_{kk'})}{\sum_{k=1}^p \sum_{k'=1}^p \text{Var}(t_{kk'} - s_{kk'}) + (\text{E}(t_{kk'} - s_{kk'}))^2} \quad (3.13)$$

Since R'' given as:

$$R''(\lambda) = 2 \sum_{k=1}^p \sum_{k'=1}^p (\text{Var}(t_{kk'} - s_{kk'}) + (\text{E}(t_{kk'} - s_{kk'}))^2),$$

is positive everywhere, the solution given in Equation (3.13), is verified as a minimum of the risk function (Ledoit and Wolf, 2003). The estimated optimal shrinkage intensity is obtained by replacing the parameters (i.e., $t_{kk'}$, $s_{kk'}$, and $\sigma_{kk'}$) in Equation (3.13) by their unbiased estimates.

$$\hat{\lambda}^* = \frac{\sum_{k=1}^p \sum_{k'=1}^p \widehat{\text{Var}}(s_{kk'}) - \widehat{\text{Cov}}(t_{kk'}, s_{kk'})}{\sum_{k=1}^p \sum_{k'=1}^p \text{E}(t_{kk'} - s_{kk'})^2}. \quad (3.14)$$

Selecting a suitable target covariance matrix \mathbf{T} is an important step in the estimation of the shrinkage estimate for the covariance matrix given in Equation (3.9). The choice of a target should be guided by the presumed lower-dimensional structure in the data set as this determines the increase in efficiency over that of the empirical covariance estimator (Schäfer and Strimmer, 2005). Several researchers have proposed various target matrices, \mathbf{T} , and every target matrix has a different variance-bias trade-off with respect to the unknown covariance matrix; thus, there is no ideal target for reducing the MSE (Lancewicki and Aladjem, 2014). Here, we consider the following four potential \mathbf{T} matrices (see Table 3.1) that are commonly used in the literature (Schäfer and Strimmer, 2005). The target matrix, \mathbf{T}_1 , is the perfect positive correlation model given by Ledoit and Wolf (2003). \mathbf{T}_2 is the constant correlation model given by Ledoit and Wolf (2004b). The

target matrices \mathbf{T}_3 and \mathbf{T}_4 shrink the diagonal of the sample covariance matrix to the mean Ledoit and Wolf (2004a) and median Opgen-rhein and Strimmer (2007) of the sample variances across all variables, respectively. See Table 3.1 for the resulting optimal shrinkage intensities for each of the four target matrices. The parameter $\Phi_{kk'}$ in Equations (3.15) and (3.16), is given as (Schäfer and

Table 3.1: Table of target matrices and their optimal shrinkage intensities. The terms $\text{avg}(\cdot)$ and $\text{Med}(\cdot)$ (in Equations (3.17) and (3.18)) are used to denote the mean and median of the sample variances across all variables, respectively.

Target matrix (\mathbf{T})	Expression	Optimal shrinkage intensity
\mathbf{T}_1	$t_{kk'} = \begin{cases} \sigma_{kk}, & \text{if } k = k' \\ \sqrt{\sigma_{kk}\sigma_{k'k'}}, & \text{if } k \neq k' \end{cases}$	$\hat{\lambda}^* = \frac{\sum_{k \neq k'} \widehat{\text{Var}}(\sigma_{kk'}) - \Phi_{kk'}}{\sum_{k \neq k'} (\sigma_{kk'} - \sqrt{\sigma_{kk}\sigma_{k'k'}})^2} \quad (3.15)$
\mathbf{T}_2	$t_{kk'} = \begin{cases} \sigma_{kk}, & \text{if } k = k' \\ \bar{\rho}\sqrt{\sigma_{kk}\sigma_{k'k'}}, & \text{if } k \neq k' \end{cases}$	$\hat{\lambda}^* = \frac{\sum_{k \neq k'} \widehat{\text{Var}}(\sigma_{kk'}) - \bar{\rho}\Phi_{kk'}}{\sum_{k \neq k'} (\sigma_{kk'} - \bar{\rho}\sqrt{\sigma_{kk}\sigma_{k'k'}})^2} \quad (3.16)$
\mathbf{T}_3	$t_{kk'} = \begin{cases} \text{avg}(\sigma_{kk}), & \text{if } k = k' \\ 0, & \text{if } k \neq k' \end{cases}$	$\hat{\lambda}^* = \frac{\sum_{k \neq k'} \widehat{\text{Var}}(\sigma_{kk'}) + \sum_{kk} \widehat{\text{Var}}(\sigma_{kk})}{\sum_{k \neq k'} (\sigma_{kk'}^2 + \sum_{kk} (\sigma_{kk} - \text{avg}(\sigma_{kk}))^2)} \quad (3.17)$
\mathbf{T}_4	$t_{kk'} = \begin{cases} \text{Med}(\sigma_{kk}), & \text{if } k = k' \\ 0, & \text{if } k \neq k' \end{cases}$	$\hat{\lambda}^* = \frac{\sum_{k \neq k'} \widehat{\text{Var}}(\sigma_{kk'})}{\sum_{k \neq k'} (\sigma_{kk'}^2 - \text{Med}(\sigma_{kk}))^2} \quad (3.18)$

Strimmer, 2005):

$$\Phi_{kk'} = \frac{1}{2} \left\{ \sqrt{\frac{\sigma_{kk}}{\sigma_{k'k'}}} \widehat{\text{Cov}}(\sigma_{kk}, \sigma_{kk'}) + \sqrt{\frac{\sigma_{k'k'}}{\sigma_{kk}}} \widehat{\text{Cov}}(\sigma_{k'k'}, \sigma_{kk'}) \right\}$$

We refer the reader to Schäfer and Strimmer (2005); Ledoit and Wolf (2003, 2004b,a); Opgen-rhein and Strimmer (2007); Ardia, David and Boudt, Kris and Gagnon Fleury (2017) for more detailed descriptions of target matrices.

3.4 Simulation study of the performance of MEWMA control charts

3.4.1 Methods

We compared the in-control and out-of-control performance of individual-observation ($n = 1$) MEWMA control charts based on each of three methods to estimate the variance-covariance matrix, described in Section 3.3.2, using simulated data. Methods were compared using the Average Run Length (ARL) before an out-of-control signal was produced.

For each simulation, the Phase I population consisted of a p -variate normal distribution with mean vector, $\boldsymbol{\mu}_0$, and covariance matrix, $\boldsymbol{\Sigma}_0$. From Phase I to Phase II, a shift of size δ (as defined in Equation 3.4) was applied to $\boldsymbol{\mu}_0$. Note, to estimate the in-control Average Run Length (ARL₀), a null shift (i.e., $\delta = 0$) was applied to simulate no difference in population parameters between Phase I and Phase II.

For each set of simulations, the following procedure was repeated 50,000 times:

- Phase I
 - (i) Generate a sample of $i = 1, \dots, m$ Phase I baseline observations, \mathbf{x}_i , each a vector of length p , from the in-control population, where m is the number of Phase I time points.
 - (ii) Estimate the covariance matrix ($\boldsymbol{\Sigma}_0$) and location ($\boldsymbol{\mu}_0$) parameters from the Phase I sample, and use these to estimate the MEWMA control chart statistic \hat{T}^2 in Phase II.
- Phase II
 - (i) Set $j = 1$, where j indexes the Phase II time points.

- (ii) Generate a random vector \mathbf{x}_j from the Phase II population and adjoin it to samples from the previous time points, \mathbf{x}_{j-1} . Estimate the MEWMA control chart statistic \hat{T}_i^2 .
- (iii) If $\hat{T}_j^2 > h*$ (the upper control limit), stop and record $j - 1$ as the Run Length for this simulation, indicating the time point before the control chart statistic first signals an out-of-control state. If $\hat{T}^2 \leq h*$, update $j = j + 1$ and return to step (ii).

At the conclusion of 50,000 simulations, the average run length (ARL) was obtained by taking the average across the recorded values of RL .

Without loss of generality, we used $\boldsymbol{\mu}_0 = \mathbf{0}$ and considered covariance matrix, $\boldsymbol{\Sigma}_0$, with variances given as:

$$\sigma_{kk'} = \begin{cases} 1, & \text{if } k = k' \\ b, & \text{if } k \neq k'. \end{cases} \quad (3.19)$$

We examined the performance of the methods under both independent and AR(1) covariance structures, by setting $b = 0$ (i.e., independent covariance structure), and $b = 0.8^{|k-k'|}$ for $1 \leq k, k' \leq p$ (i.e., AR(1) covariance structure), respectively. Our simulations included Phase I sample sizes of $m = 30, 40, 50, 70, 100, 200, 300$, and 500.

For each simulation, the values of the upper control limit h were chosen such that the chart would produce an ARL_0 of 200 for known $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. For establishing control limits from Phase I, we used three different values of the smoothing parameter: $r = 0.05, 0.1$, or 0.2 . The MEWMA chart is known to work well with these choices of r , and they are popular choices in practice (Montgomery, 2009).

A common approach to reducing the frequency of false out-of-control signals of control charts for which the population parameters are estimated is to use wider, “corrected” control limits, h_c

(Aly et al., 2015). See Table 3.2 for the pairs of the values of r and h that fixed ARL_0 to 200 for the known parameter case, and Section 3.4.2 for correcting control limits that give ARL_0 of 200 (approximately) when population parameters are unknown. A binary search algorithm similar to the one given in Champ et al. (2005) and Mahmouda and Maravelakisb (2010) was used to obtain the corrected limits (h_c). We refer to the results obtained using target matrices \mathbf{T}_1 , \mathbf{T}_2 , \mathbf{T}_3 , or \mathbf{T}_4 , as shrinkage(T_1), shrinkage(T_2), shrinkage(T_3), and shrinkage(T_4), respectively.

Table 3.2: The upper control limit, h , that produces $AARL_0$ of 200 when the MEWMA chart statistic is obtained from known parameters, for different values of p and r .

p	r		
	0.05	0.1	0.2
2	7.36	8.67	9.67
3	9.41	10.77	11.87
4	11.21	12.72	13.86
5	12.97	14.54	15.75

3.4.2 Results of the simulation study

In-control performance

The in-control performance is shown in plots of the (ARL_0) for simulated MEWMA charts under the independent covariance structure and the AR(1) covariance structure, respectively (Figures 3.1 and 3.2), when $r = 0.05$, for $p = 2, 3, 4$, or 5 and different values for m . In all cases, the ARL_0 of the MEWMA charts was substantially less than the benchmark of 200, which would be achieved if the known parameters were used to establish the control limits. This demonstrates how estimating the population parameters compromises the performance of the chart by increasing the frequency of false out-of-control signals. The ARL_0 performance of the chart depended on the parameters used (m , and p) and the method used to estimate the covariance structure. The shrinkage(T_4) and

MSSD covariance estimates had the highest and lowest ARL_0 values, respectively. Unsurprisingly, as the value of m increased, the (ARL_0) value increased, regardless of the method used to estimate Σ_0 . In particular, the ARL_0 of all the methods approached the nominal ARL_0 value (200) for large values of m . Tables 3.3 and 3.4 give the ARL_0 values of the MEWMA charts under the independent and AR(1) covariance structures, respectively, when $r = 0.1$.

For every parameter combination we considered, the ARL_0 performance was best for the shrinkage(T_4) method, and the improvement of the shrinkage(T_4) method was most evident with small m and higher p . In these scenarios, the shrinkage(T_4) method ARL_0 values were up to three times as high as the worst estimator (see Tables 3.3 and 3.4, or Figures 3.1 and 3.2). When $p=2$, the shrinkage(T_3) method consistently had the lowest ARL_0 , but for $p > 2$, the MSSD estimator typically had the lowest ARL_0 . The performance of the MEWMA charts based on the empirical estimate, and MSSD covariance estimates were consistent in terms of ARL_0 across both covariance structures examined here. On the other hand, the ARL_0 performance of the MEWMA chart based on shrinkage estimates varied across the two different covariance structures, with higher ARL_0 values obtained for the AR(1) covariance structure compared to those obtained for the identity covariance matrix.

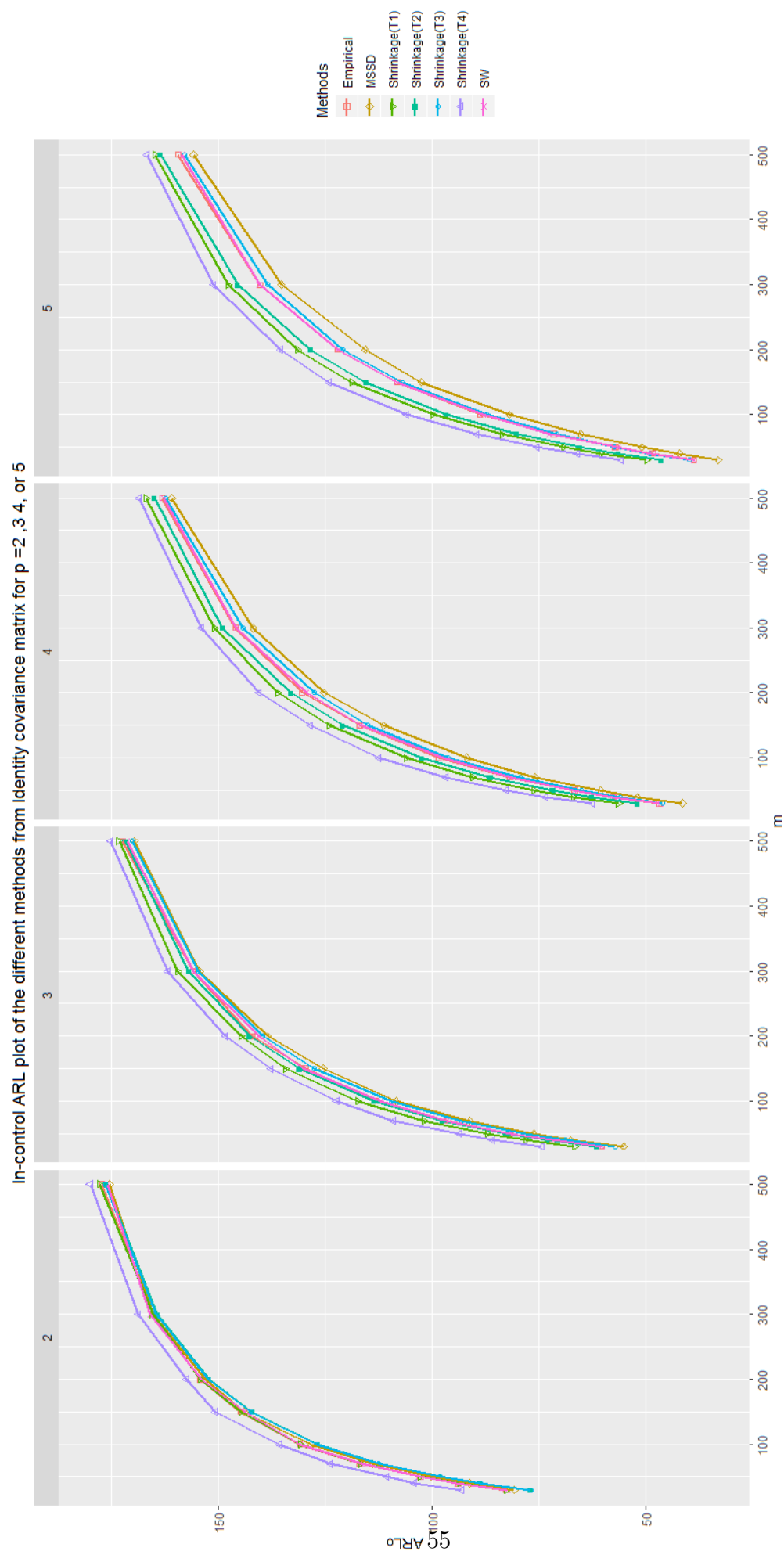


Figure 3.1: In-control $AARL$ plot from identity covariance matrix for $r = 0.05$, and different values of p , and m

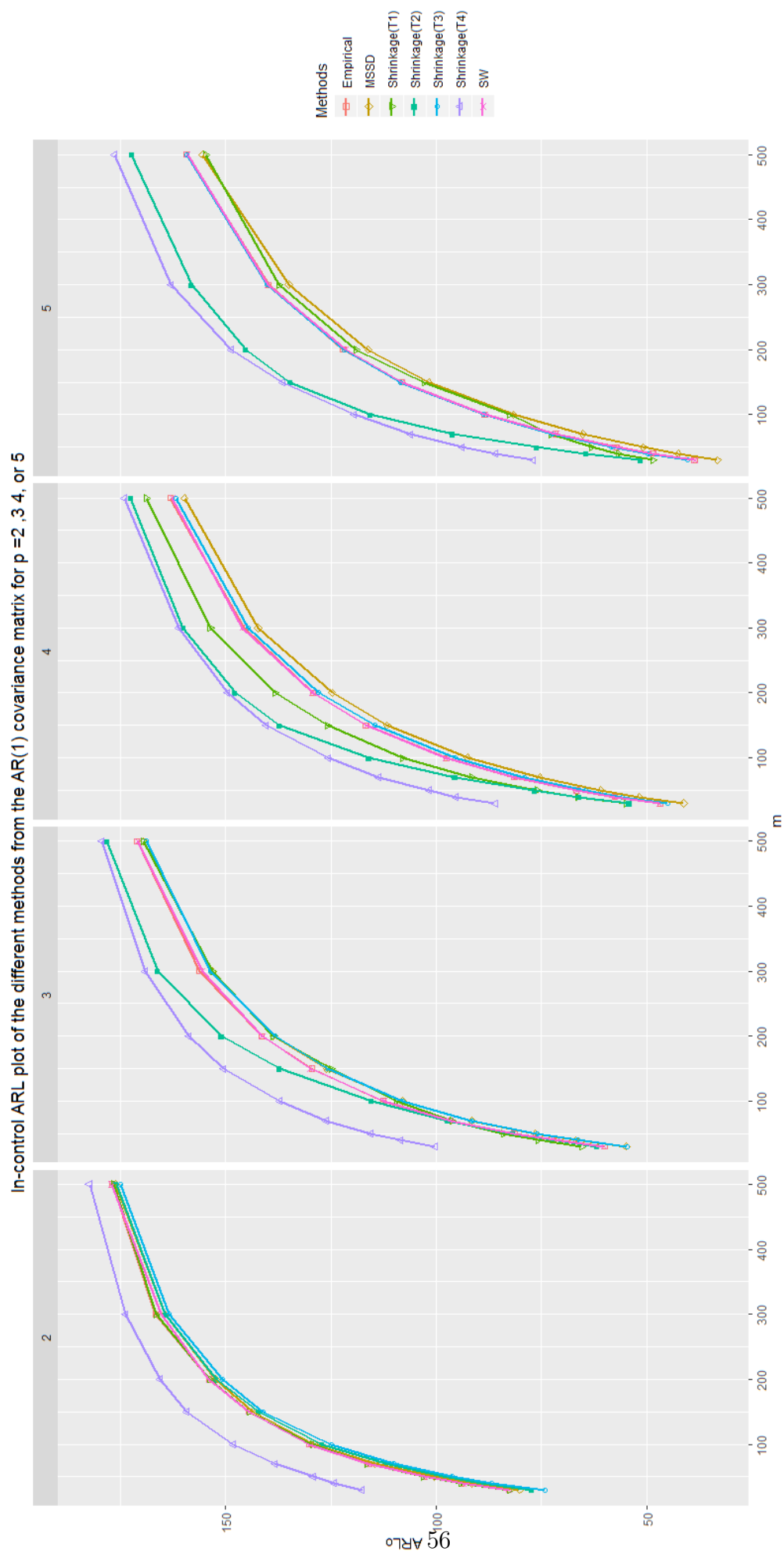


Figure 3.2: In-control $AARL$ plot from the AR(1) covariance structure for $r = 0.05$, and different values of p , and m

Table 3.3: $AARL_0$ values when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters under independent covariance structure. The $AARL_0$ is fixed at 200, and $r = 0.1$.

		m								
p	$Method$	30	40	50	70	100	150	200	300	500
2	<i>Empirical</i>	92.47	106.35	114.29	127.74	141.78	155.08	164.27	174.49	183.95
	<i>MSSD</i>	90.96	100.6	111	124.8	139.39	154.57	163.33	174.83	183.59
	<i>Shrinkage(T1)</i>	93.32	104.26	113.44	127.15	141.17	154.88	164.51	173.78	183.9
	<i>Shrinkage(T2)</i>	84.73	97.14	107.19	122.3	136.96	151.63	161.33	172.82	182.97
	<i>Shrinkage(T3)</i>	84.12	97.33	106.89	122.17	137.05	151.9	161.52	171.92	183.55
	<i>Shrinkage(T4)</i>	111.67	121.24	128.35	139.87	150.51	162.6	169.59	178.65	187.94
3	<i>Empirical</i>	62.38	74.36	85.57	101.94	117.69	136.74	147.42	160.2	173.25
	<i>MSSD</i>	55.52	67.47	79.26	95.66	112.33	131.18	142.34	157.49	170.26
	<i>Shrinkage(T1)</i>	73.08	85.47	95.77	110.46	126.14	142.13	152.77	164.38	175.98
	<i>Shrinkage(T2)</i>	65.16	77.91	88.76	104.8	121.27	137.96	148.17	162.75	174.75
	<i>Shrinkage(T3)</i>	58.14	71.24	81.61	97.28	114.44	133.01	144.86	158.93	172.11
	<i>Shrinkage(T4)</i>	87.98	99.92	108.3	122.18	135.94	149.93	158.61	169.5	179.77
4	<i>Empirical</i>	46.37	58.55	69.59	85.5	102.85	124.51	136.14	153.36	169.74
	<i>MSSD</i>	39.2	50.84	60.72	77.61	96.74	115.93	130.53	147.38	163.82
	<i>Shrinkage(T1)</i>	62.2	74.69	84.96	100.82	117.09	135.12	146.46	160.19	173.95
	<i>Shrinkage(T2)</i>	55.24	67.7	78.06	94.79	112.15	130.7	143.21	157.81	171.42
	<i>Shrinkage(T3)</i>	45.36	57.07	67.07	83.02	101.19	120.96	134.39	150.72	167.24
	<i>Shrinkage(T4)</i>	74.6	86.95	96.91	111.24	126.68	143.39	152.71	165.51	176.55
5	<i>Empirical</i>	36.5	47.2	56.91	73.64	91.62	112.9	125.54	144.45	162.04
	<i>MSSD</i>	29.31	38.97	48.4	63.7	81.62	104	117.81	136.3	158.18

<i>Shrinkage(T1)</i>	54.71	66.81	76.78	93.26	110.63	129.65	141.63	157.32	171.71
<i>Shrinkage(T2)</i>	48.55	60.56	70.97	87.5	104.9	125.53	137.63	154.67	168.98
<i>Shrinkage(T3)</i>	38.34	48.39	57.54	72.58	90.69	111.66	126.11	143.91	162.01
<i>Shrinkage(T4)</i>	66.28	78.08	88.11	104.16	120.11	137.35	148.54	162.35	175.32

Table 3.4: $AARL_0$ values when different Phase I samples (m), of size $n = 1$ are used in estimating the unknown parameters under AR(1) covariance structure. The $AARL_0$ is fixed at 200, and $r = 0.1$.

p	$Method$	m								
		30	40	50	70	100	150	200	300	500
2	<i>Empirical</i>	92.42	104.43	114.31	127.54	141.4	155.86	164.2	174.96	183.48
	<i>MSSD</i>	90.19	102.63	111.8	125.65	139.63	153.77	162.81	173.13	183.28
	<i>Shrinkage(T1)</i>	92.79	105.04	113.84	127.02	141.13	154.95	164.17	174.23	183.98
	<i>Shrinkage(T2)</i>	84.1	97.34	106.68	121.97	136.83	151.76	161.56	172.44	182.93
	<i>Shrinkage(T3)</i>	79.61	92.83	103.48	119.28	134.61	149.72	160.23	170.81	182.31
	<i>Shrinkage(T4)</i>	153.8	154.95	157.45	162.57	167.94	175.43	180.03	185.49	190.99
3	<i>Empirical</i>	62.22	75.89	85.71	101.93	118.62	135.74	147.09	161.01	173.61
	<i>MSSD</i>	55.7	68.48	79.17	95.27	112.35	131.12	142.97	157.63	171.53
	<i>Shrinkage(T1)</i>	68.07	78.28	86.11	98.28	109.93	121.44	129.05	137.52	145.16
	<i>Shrinkage(T2)</i>	63.04	75.06	84.55	100.86	120.17	143.68	158.25	173.98	184.98
	<i>Shrinkage(T3)</i>	55.01	67.59	78.28	95.28	112.77	131.02	143.28	157.56	171.92
	<i>Shrinkage(T4)</i>	134.2	137.93	141.38	148.62	156.72	165.09	171.21	178.55	185.34
4	<i>Empirical</i>	46.87	59.1	69.23	85.64	104.13	123.94	136.82	152.85	168.65
	<i>MSSD</i>	39.09	50.77	60.71	77.3	95.58	116.47	130.2	147.06	164.71
	<i>Shrinkage(T1)</i>	54.96	63.19	69.71	79.33	88.78	98.56	103.19	110.13	116.25

	<i>Shrinkage</i> (T_2)	55.89	69.53	81.9	104.47	129.2	152.7	163.96	176.68	186.6
	<i>Shrinkage</i> (T_3)	44.55	56.28	66.4	82.42	101.07	121.3	134.2	151.14	167.89
	<i>Shrinkage</i> (T_4)	117.51	123.95	128.81	137.6	147.93	159.35	166.5	174.96	182.65
5	<i>Empirical</i>	36.51	47.45	57.35	73.49	92	113.5	127	145.73	163.38
	<i>MSSD</i>	29.38	39.34	48.46	64.19	82.5	103.83	119.09	138.31	158.7
	<i>Shrinkage</i> (T_1)	46.51	53.72	59.2	67.23	75.13	82.57	86.67	92.04	96.6
	<i>Shrinkage</i> (T_2)	53.5	69.68	84.27	108.58	132.36	153.59	164.98	176.99	185.75
	<i>Shrinkage</i> (T_3)	38.67	49.63	58.9	74.94	93.2	113.93	127.7	145.22	162.96
	<i>Shrinkage</i> (T_4)	106.46	113.19	119.56	129.41	140.71	154.17	161.77	171.3	181.08

The shrinkage(T_4) method of estimating the covariance matrix consistently outperformed the alternatives for the in-control simulations examined here. Thus, we restrict our attention to the shrinkage(T_4) method for the remainder of this paper, and henceforth refer to it as the “shrinkage covariance estimate”.

Corrected limits

As shown in Table 3.5, the corrected upper control limits (h_c) depend on the Phase I sample size, m , smoothing parameter, r , and the number of quality characteristics being monitored, p . Smaller values of m give bigger values of h_c , and as m becomes larger, the values of h_c in all cases approach the corresponding values of h when the parameters are known (given in Table 3.2). Our simulation results show that the corrected values h_c for the chart based on the shrinkage-MEWMA method were much smaller than those based on other methods used to estimate the covariance matrix. In section 3.4.2, we used the values of h_c obtained here to measure and compare the out-of-control performance of the MEWMA control charts using simulations.

Table 3.5: The corrected upper control limit (h_c), that produced $AARL_0$ of 200 when different Phase I samples (m), of size $n = 1$ are used to estimate the unknown parameters.

			m								
r	p	<i>Method</i>	30	40	50	70	100	150	200	300	500
0.05	2	<i>Empirical</i>	10.3045	9.802	9.4882	9.0669	8.6452	8.2891	8.1037	7.8911	7.6796
		<i>MSSD</i>	10.331	9.8457	9.5528	9.0807	8.6884	8.3135	8.1286	7.8742	7.6834
		<i>Shrinkage</i>	9.6936	9.3792	9.135	8.7997	8.4731	8.1793	8.0089	7.8216	7.6638
	3	<i>Empirical</i>	14.8071	13.7325	13.0491	12.2132	11.4761	10.854	10.5344	10.2145	9.8794
		<i>MSSD</i>	15.3062	14.1393	13.3982	12.4574	11.6538	10.9725	10.6354	10.2375	9.919
		<i>Shrinkage</i>	13.1039	12.5825	12.1187	11.5759	11.0457	10.5956	10.3357	10.0421	9.7923
	4	<i>Empirical</i>	19.4159	17.6778	16.5469	15.164	14.202	13.2499	12.7834	12.2821	11.8679
		<i>MSSD</i>	20.7546	18.6235	17.2884	15.7353	14.472	13.4902	12.9557	12.397	11.9454
		<i>Shrinkage</i>	16.4152	15.5326	14.9511	14.1624	13.4278	12.8067	12.4389	12.0815	11.7468
	5	<i>Empirical</i>	24.2795	21.6406	20.0546	18.175	16.7421	15.5848	14.9614	14.2996	13.7754
		<i>MSSD</i>	26.663	23.3208	21.2897	19.018	17.257	15.93	15.1996	14.4738	13.8642
		<i>Shrinkage</i>	19.4002	18.3263	17.5732	16.5382	15.6169	14.8838	14.4205	13.9927	13.6046
0.1	2	<i>Empirical</i>	10.8482	10.4845	10.2244	9.8796	9.5922	9.2919	9.1673	8.9986	8.8688
		<i>MSSD</i>	10.8606	10.4994	10.2522	9.92	9.6283	9.3244	9.1816	9.014	8.8862
		<i>Shrinkage</i>	10.1993	9.9722	9.8344	9.614	9.3926	9.1959	9.0686	8.944	8.8279
	3	<i>Empirical</i>	14.9746	14.2128	13.64	12.9718	12.3835	11.8999	11.6567	11.3997	11.1584
		<i>MSSD</i>	15.6048	14.6293	14.0138	13.2095	12.559	12.0309	11.7482	11.4381	11.187
		<i>Shrinkage</i>	13.2097	12.8836	12.6017	12.2196	11.8971	11.5441	11.427	11.2052	11.0673
	4	<i>Empirical</i>	19.3746	17.8772	16.9523	15.8482	14.9785	14.2791	13.9232	13.5324	13.232
		<i>MSSD</i>	20.6208	18.8435	17.7037	16.3796	15.3575	14.5307	14.1056	13.659	13.2747
		<i>Shrinkage</i>	16.1222	15.5268	15.1761	14.6301	14.1421	13.7581	13.5238	13.2634	13.0532

0.2	5	<i>Empirical</i>	23.855	21.5848	20.2327	18.7175	17.495	16.5624	16.0867	15.569	15.1711
		<i>MSSD</i>	26.1712	23.2595	21.5371	19.5862	18.1059	16.9168	16.3738	15.7235	15.285
		<i>Shrinkage</i>	18.7201	17.97	17.497	16.8198	16.2627	15.7648	15.4659	15.1851	14.9572
	2	<i>Empirical</i>	11.1606	10.9116	10.7398	10.5191	10.274	10.1044	10.0032	9.9033	9.811
		<i>MSSD</i>	11.0874	10.9242	10.7311	10.5328	10.3165	10.1194	10.016	9.9133	9.8058
		<i>Shrinkage</i>	10.4287	10.3257	10.2948	10.1379	10.074	9.9424	9.8857	9.7967	9.7598
	3	<i>Empirical</i>	15.1319	14.4709	14.0211	13.495	13.0492	12.7005	12.5097	12.2801	12.1339
		<i>MSSD</i>	15.7713	14.9017	14.38	13.7495	13.2664	12.8158	12.5937	12.3468	12.1493
		<i>Shrinkage</i>	13.2171	13.0026	12.8544	12.6789	12.4821	12.3081	12.2149	12.1013	12.0143
	4	<i>Empirical</i>	19.2106	17.9314	17.1904	16.3061	15.5966	15.0579	14.7689	14.4689	14.2336
		<i>MSSD</i>	20.5207	18.9568	18.0405	16.8761	16.0205	15.3038	14.9591	14.6297	14.319
		<i>Shrinkage</i>	15.8131	15.5067	15.2599	14.9462	14.68	14.4334	14.3041	14.174	14.0642
	5	<i>Empirical</i>	23.3851	21.5063	20.3793	19.0875	18.0828	17.294	16.9229	16.5205	16.214
		<i>MSSD</i>	25.8751	23.2379	21.7234	19.994	18.7472	17.7318	17.2322	16.7384	16.3304
		<i>Shrinkage</i>	18.1071	17.7156	17.3851	16.9978	16.6953	16.4062	16.2625	16.0835	15.9355

We provide *R-code* (R Core Team, 2013) (in the supplementary material) which practitioners can use to obtain corrected limits, (h_c) , for MEWMA charts based on different covariance estimators, different structures for Σ_0 , or for Phase I sample sizes (m) that differ from those considered in this study. Also, practitioners can use the *R-code* to obtain corrected limits (h_c) for MEWMA charts that fix the in-control ARL at a value different from the one considered in this study (i.e., $ARL_0 = 200$). We adopted the ARL numerics algorithm for a MEWMA chart Knoth (2017), implemented in the package *spc* Knoth (2018), to obtain an arbitrary start value (say h'_c), and used a binary search algorithm to search for the corrected limit.

Out-of-control performance using corrected limits

The corrected limits given in Section 3.4.2 allowed us to standardize the in-control performance to $ARL_0 = 200$ for each of the methods. However, when estimating parameters, there will be an increase in the mean out-of-control ARL compared with charts that use *a priori* known process parameters. In this section, we compare the performance of control charts, using each of the methods of estimating Σ_0 from Phase I data, with respect to detecting an out-of-control process in Phase II. The simulation procedure is outlined in Section 3.4.2, with the process mean vector shifted between Phases I and II by values of size $\delta = 0.5, 1, 1.5, 2, 2.5$ or 3 .

We used a modification of the extra quadratic loss (EQL) to measure performance over a range of shift values (Wu et al., 2009; Ahmad et al., 2013; Adegoke et al., 2018b). EQL is the weighted average of the ARL over the entire set of shifts observed across the charting structure, using δ^2 as the weights (Ahmad et al., 2013). The modified EQL (MEQL) performance of each control chart was evaluated by first calculating ARL_δ (i.e., ARL of the chart based on the different methods), for each set of simulation. We then took the difference between the ARL_δ for each method with $ARLB_\delta$ (the ARL of a benchmark control chart based on known Phase I population parameters μ_0 and Σ_0). These differences were then converted into an overall measure of performance by calculating the extra quadratic loss, weighted by δ^2 , of the ARL scores across values of δ . This Modified Extra Quadratic Loss is given by:

$$MEQL = \frac{1}{\delta_{max} - \delta_{min}} \int_{\delta_{min}}^{\delta_{max}} \delta^2 (ARL_\delta - ARLB_\delta) d(\delta). \quad (3.20)$$

where δ^2 is the square of the shift size δ , and δ_{max} and δ_{min} are the maximum and minimum shifts, respectively. The integral was solved numerically. A chart with a very low MEQL is, on average, considered to perform almost as well as the benchmark chart.

Of the different methods used to estimate Σ_0 , the shrinkage(T_4) method consistently gave the best

Table 3.6: The Modified Average Quality Loss (MAQL) for control charts using various methods of estimating the covariance matrix. Results are shown for dimensionality $p = 2, 3, 4$ or 5 and smoothing parameter $r = 0.05$. The corrected limits presented in Table 3.5, which standardized the $AARL_0$ to 200, were used.

p	$Methods$	m								
		30	40	50	70	100	150	200	300	500
2	<i>Empirical</i>	3.79	3.12	2.68	2.10	1.62	1.18	0.99	0.69	0.44
	<i>MSSD</i>	3.53	2.97	2.61	2.06	1.60	1.18	0.94	0.72	0.45
	<i>Shrinkage</i>	3.50	2.93	2.46	1.93	1.53	1.15	0.90	0.66	0.42
3	<i>Empirical</i>	5.36	4.32	3.67	2.83	2.13	1.55	1.22	0.86	0.55
	<i>MSSD</i>	5.18	4.29	3.73	2.87	2.23	1.58	1.25	0.86	0.54
	<i>Shrinkage</i>	4.40	3.70	3.20	2.49	1.86	1.38	1.09	0.74	0.45
4	<i>Empirical</i>	6.55	5.47	4.59	3.53	2.65	1.94	1.50	1.08	0.67
	<i>MSSD</i>	6.55	5.52	4.69	3.72	2.78	1.97	1.56	1.09	0.69
	<i>Shrinkage</i>	5.53	4.50	3.94	3.03	2.34	1.67	1.32	0.93	0.57
5	<i>Empirical</i>	7.86	6.42	5.38	4.09	3.07	2.21	1.71	1.14	0.76
	<i>MSSD</i>	7.76	6.70	5.68	4.34	3.27	2.29	1.77	1.24	0.76
	<i>Shrinkage</i>	6.16	5.10	4.40	3.41	2.57	1.90	1.47	1.01	0.62

overall out-of-control performance. The superiority of the shrinkage method was most pronounced for smaller values of m ; as m increased, the performance of all methods converged to that of the benchmark chart. (Table 3.6).

3.5 Illustrative examples

In this section, we illustrate the performance of the MEWMA chart obtained using the proposed shrinkage estimate and the other methods for estimating the covariance matrix using two different real data sets. The first dataset is a bimetal thermostat data set (5 variables); the second is a high-dimensional gene expression data set (271 variables). We applied each of the methods for estimating Σ_0 (Section 3.3.2) and compared the performance of control charts derived from them. In each analysis, μ_0 and Σ_0 were estimated from Phase I data. We used $r = 0.05$ with two different control limits: the upper control limits h that give ARL_0 of 200 when parameters are known, and the corrected upper control limits h_c that give

ARL₀ of 200 when parameters are estimated from the Phase I sample. These values (i.e., the h_c 's) were obtained using the binary search algorithm referred to above in Section 3.4.2.

3.5.1 Example 1: Bimetal thermostat dataset

Bimetal thermostats have a bimetallic strip, composed of two strips of different metals, which together convert changes in temperature to a mechanical displacement *via* thermal expansion, so that the temperature may be read (Santos-Fernandez, 2012). The dataset contains measurements of deflection, curvature, resistivity and hardness for each of the low-and-high expansion sides, from brass and steel bimetal thermostats. These data were given in Santos-Fernandez (2012), and Santos-Fernandez, Edgar and Santos-Fernandez (2016). The data are divided into two Phases, each comprised of $m = 28$ time points and $p = 5$ variables.

When the upper control limits h (given in Table 3.2) were used, the control charts using the empirical estimate of the covariance matrix or MSSD both detected the first signal after the 14th observation (see Figure 3.3). However, the control chart based on the shrinkage estimate detected the first signal after the 12th observation. When the corrected upper control limits h_c were used, the empirical estimate of the covariance matrix or MSSD failed to detect a shift. In contrast, the shrinkage method detected the first signal after the 20th observation. Here, the values of the h_c used were 27.7269, 25.0551 and 19.7351, for the MSSD, empirical and shrinkage estimates of the covariance matrix, respectively.

3.5.2 Example 2: Gene expression dataset

We also illustrate the performance of the methods on the gene expression dataset from two studies by Pawitan et al. (2005), and Miller et al. (2005). These two studies both took samples from women being treated for breast cancer in Sweden, but they span different time periods, 1987 – 1989 and 1995 – 1996, respectively. Gene expression in breast cancer tumors was measured using H133A affymetrics chips. Measurements of low quality or without associated survival data were eliminated, which yielded 232

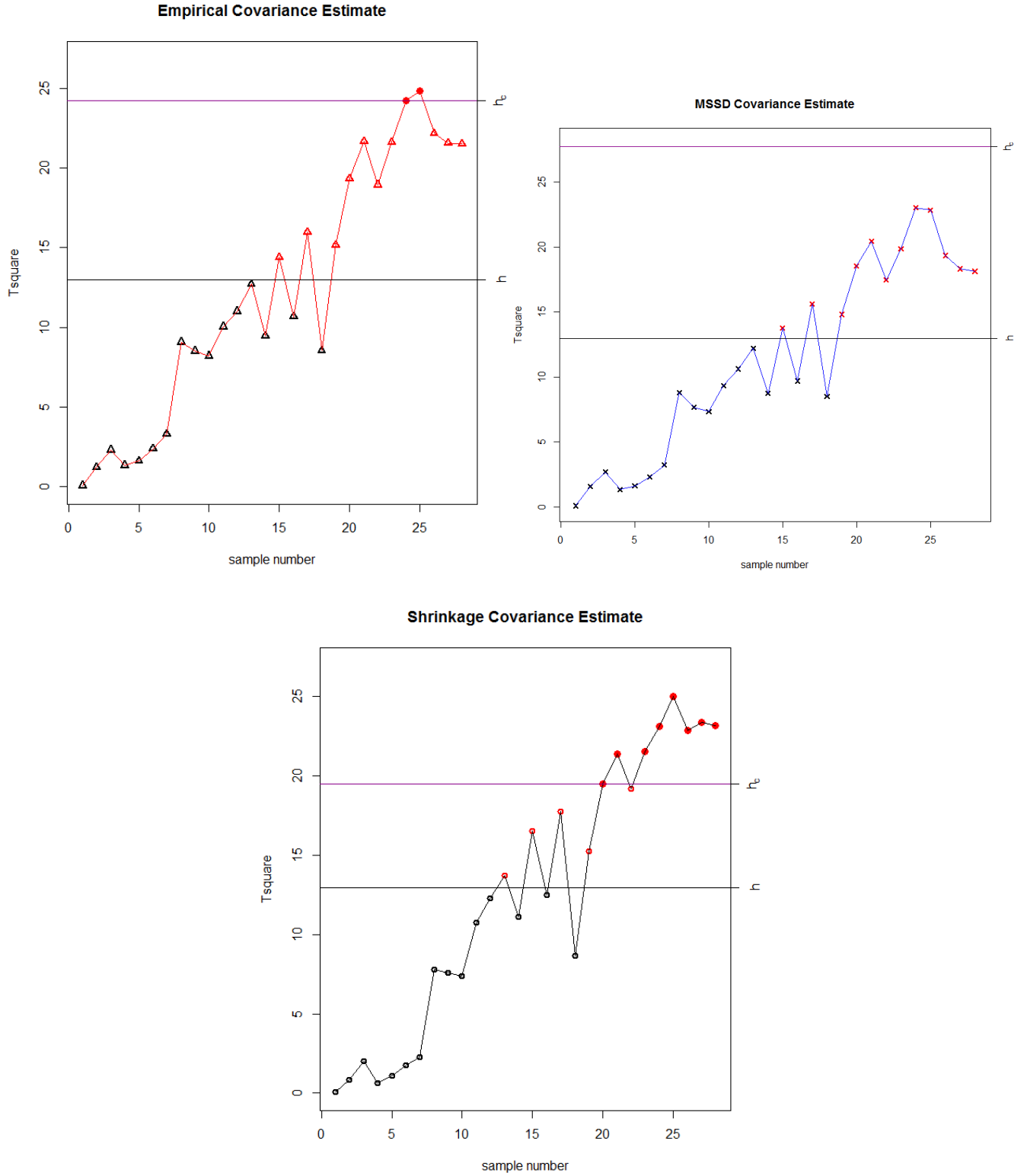


Figure 3.3: The MEWMA control charts of the Bimetal II data set. The estimated covariance matrices are obtained from the Bimetal I data set. The smoothing parameter used is $r = 0.05$ and $p = 5$. The upper control limit h , for the known parameters case; given in Table 3.2, and the corrected upper control limit h_c ; given in Table 3.5, used, are the one that give $AARL_0$ of 200.

cases for the earlier data collection period and 159 for the later one. There are some differences in the descriptions of the protocols used to collect the tissues; the earlier samples (232 patients) are described as “frozen” while the later ones (159 samples) are “frozen immediately on dry ice or in liquid nitrogen and stored in $-70^{\circ}C$ freezers”. We treated the differences between the groups as examples of differences that might arise if a protocol was altered or if a different protocol was used by mistake. Also, we used only a subset of the gene expression measurements, namely, the 271 genes associated with the *erbB-2* pathway (Thor et al., 1998). These were measured on 391 patients in total. We used the data from 1987 – 1989 as Phase I, and data from 1995 – 1996 as Phase II (Ullah, 2015).

We first reduced the dimensionality of the dataset using principal component analysis (PCA), and then constructed the control charts using the PC axes. Following the Kaiser-Guttman criterion, we retained only these PC axes whose corresponding eigenvalue, λ , was greater than one (Legendre and Legendre, 1988). Only 55 of the original $p = 271$ axes had $\lambda > 1$. Hence, we retained and used these 55 PC axes for the Phase I data set. The Phase II data set was projected onto the axes defined by these 55 PC axes of the Phase I data set. See Ferrer Ferrer (2007, 2014) for the analytical expressions for this projection. We give the MEMWA charts results for the shrinkage and empirical estimates of the covariance matrices in Figure 3.5.

When the upper control limit $h = 78.43$ was used, the control charts from both methods (i.e., the shrinkage and empirical estimates of the covariance matrix) that estimated the covariance matrix from Phase I detected the first signal after the 5th (see Figure 3.5). When the corrected upper control limits h_c were used, the empirical estimate of the covariance matrix detected out-of-control behavior after the 7th observation. In contrast, the shrinkage method detected the first signal after the 5th observation. The values of the h_c used are 122.99 and 89.12, for the empirical and shrinkage estimates of the covariance matrix, respectively.

Because the PC axes are linear combinations of the original variables, the interpretation of out-of-

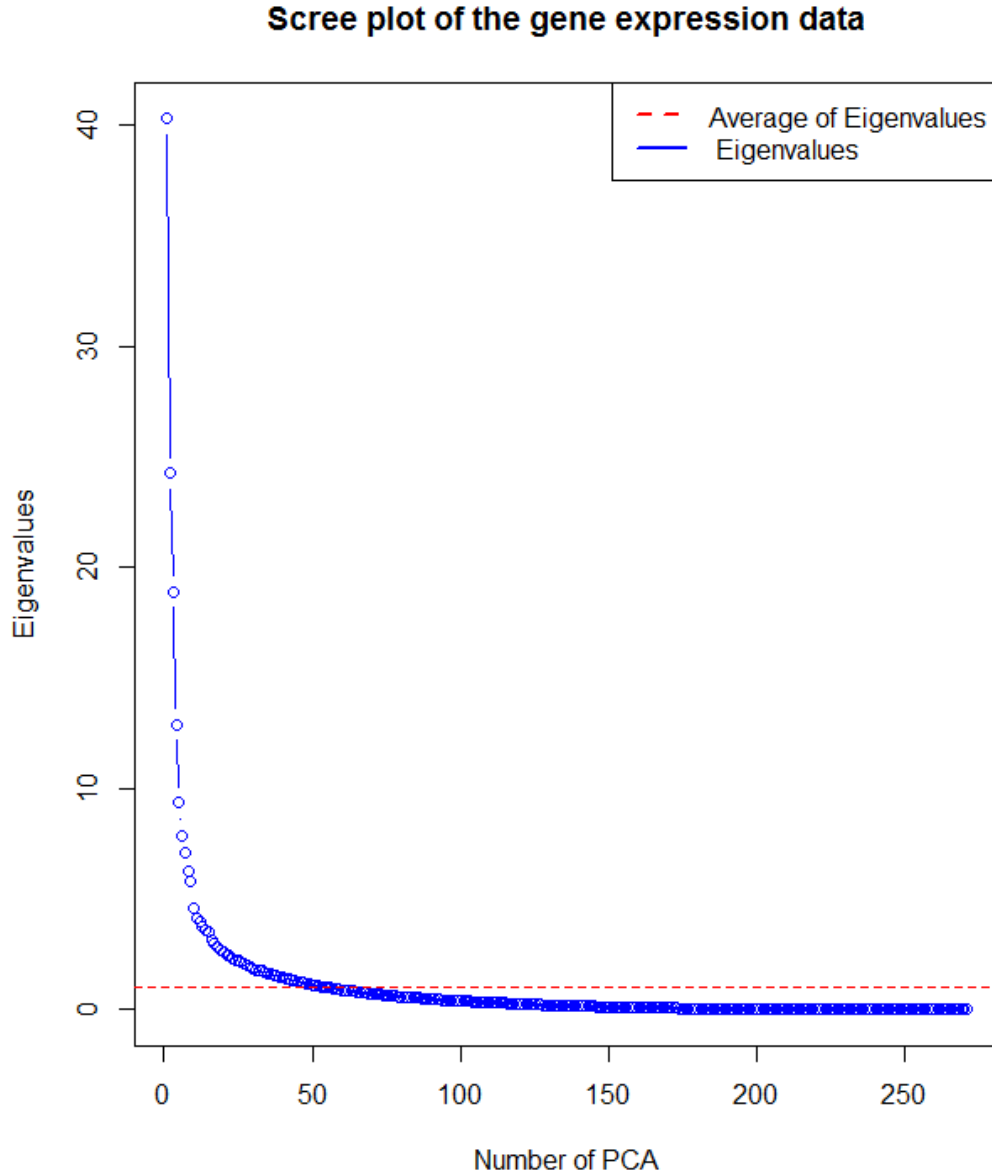


Figure 3.4: The scree plot of the gene expression data set

control signals can be difficult. Hence, we also considered control charts using the raw gene expression dataset. Due to the high-dimensionality of the Phase I data set, we were unable to calculate the inverse of the covariance matrices for either the empirical or the MSSD estimates. This was not true for the shrinkage method, however, so we can report the results of a MEWMA control chart for the Phase I and Phase II gene expression datasets using this method (Figure 3.6). All of the Phase I samples were plotted within the control-chart limit, implying that all of the Phase I samples were in-control. In Phase II, the

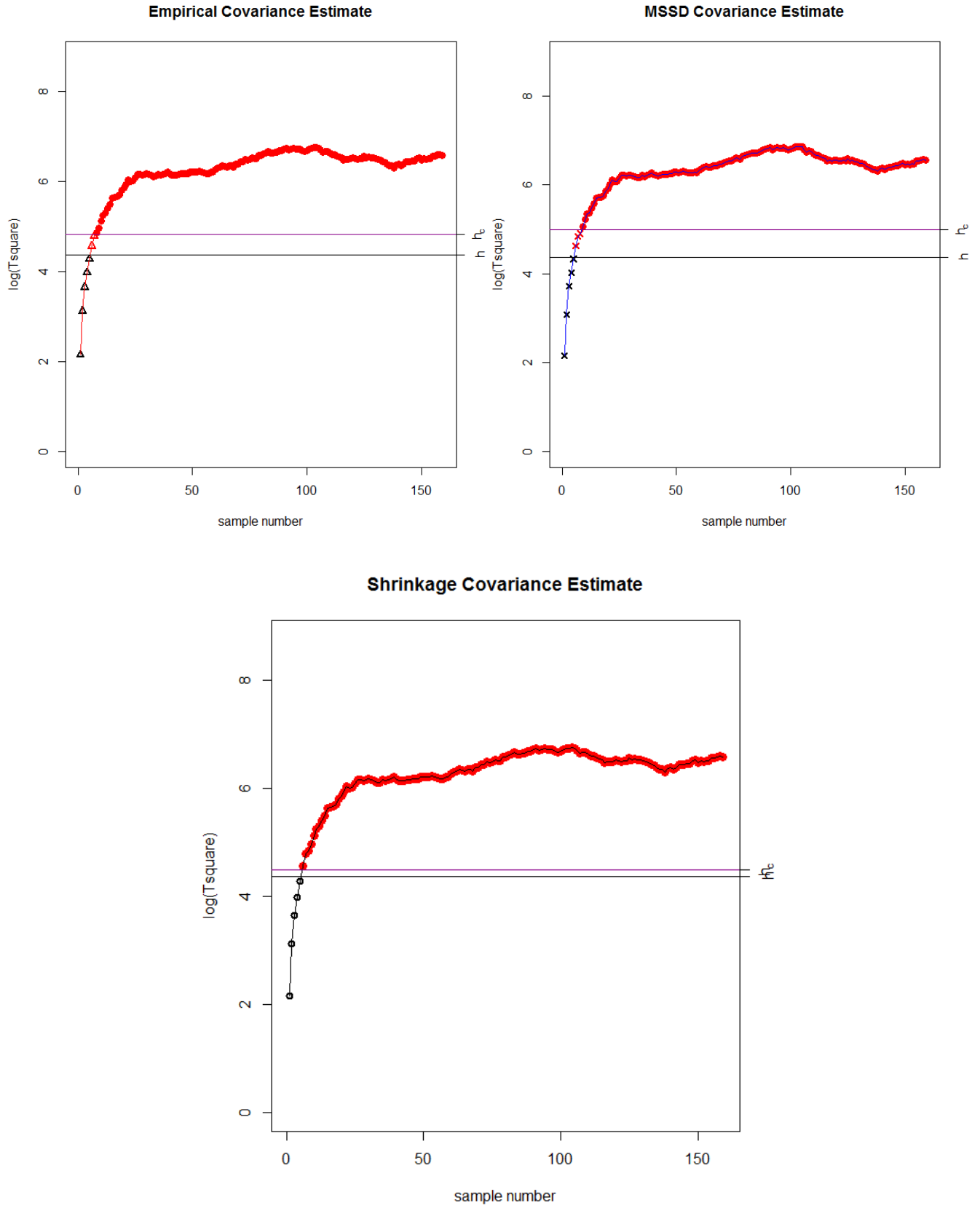


Figure 3.5: The MEWMA control charts of the gene expression data set. The estimated covariance matrices are obtained from the gene expression Phase I. The smoothing parameter used is $r = 0.05$ and the number of PC axes $k = 55$. The logarithms of upper control limit h , and the corrected upper control limit h_c are shown.

chart gave an out-of-control signal after the first sample (using $h = 323.21$ and $h_c = 369.79$).

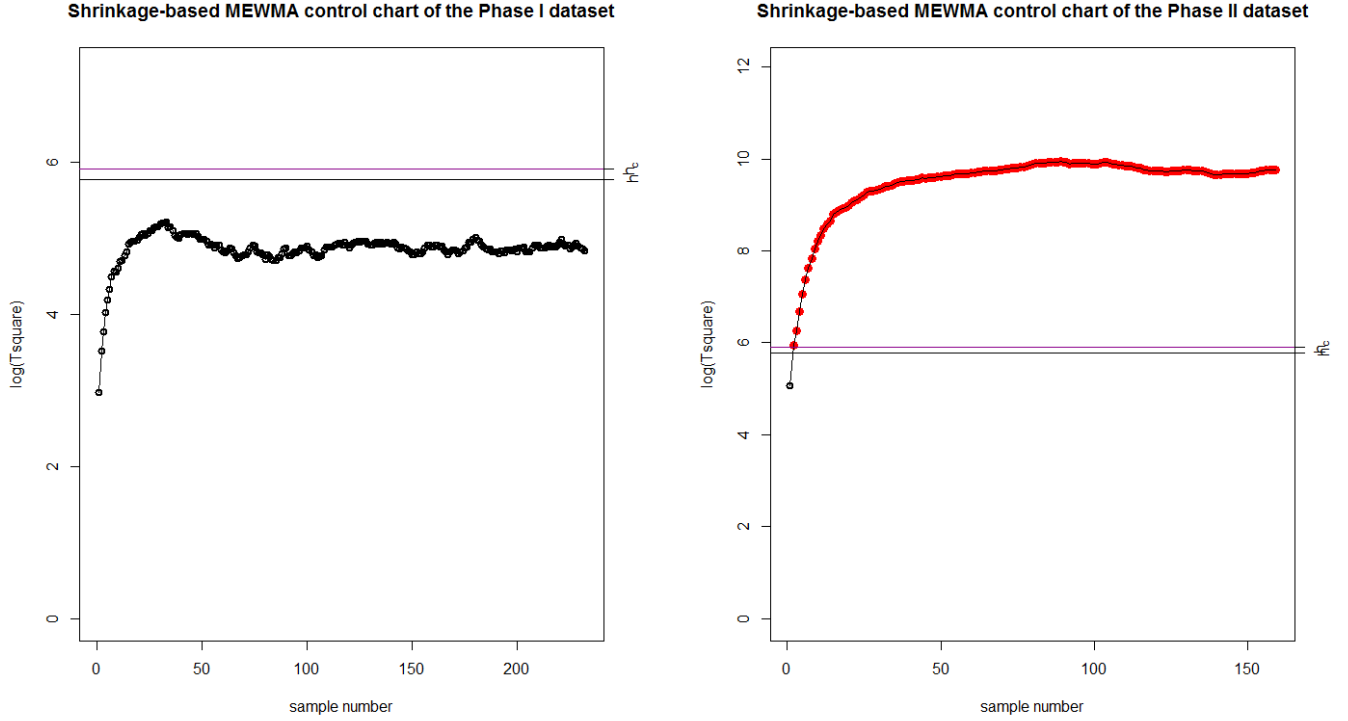


Figure 3.6: The Shrinkage-based MEWMA control chart of the gene expression for both Phase I and Phase II data set. The estimated shrinkage covariance matrices are obtained from the gene expression Phase I data set. The smoothing parameter used is $r = 0.05$ and $p = 271$. The logarithms of upper control limit h , and the corrected upper control limit h_c are shown.

3.6 General conclusions

In this paper, we studied the performance of a MEWMA control chart for individual-observation monitoring when the in-control parameters (specifically, the mean vector and covariance matrix) are estimated empirically from Phase I data, including cases where only a small Phase I sample is available. We propose the use of shrinkage methods for estimating the covariance matrix in MEWMA control charts, where the diagonal of the unbiased empirical estimate is shrunk towards its median. The performance of these proposed shrinkage methods were compared with two alternative methods that are in common usage: MEWMA control charts using either an unbiased empirical estimator or a mean square successive

difference (MSSD) estimator.

We showed, through simulation, that the in-control performance of the MEWMA control chart is strongly affected by the estimation procedure unless the Phase I sample is large. We calculated the corrected control-chart limits for different Phase I sample sizes m , such that they give a standardised ARL_0 when using estimated parameters. In simulations, MEWMA control charts obtained using the shrinkage estimate of the covariance matrix consistently performed better than other methods, especially when small Phase I samples were available from the monitoring. On average, the shrinkage estimate had longer ARL_0 for in-control processes and shorter ARL for out-of-control processes. We also demonstrated superiority of the shrinkage method with real datasets, where it detected true shifts earlier than the alternatives. The shrinkage method also has the advantage of being applicable for high-dimensional data, in contrast with the alternative methods, which cannot be used when $p > m$.

Chapter 4

Multivariate coefficient of variation (CV) control charts in Phase I of statistical process control (SPC)

”This is the peer reviewed version of the following article: “Abbasi, S. A., & Adegoke, N. A. (2018). Multivariate coefficient of variation control charts in phase I of SPC. The International Journal of Advanced Manufacturing Technology, 99(5-8), 1903-1916”, which has been published in final form at <https://doi.org/10.1007/s00170-018-2535-3>. This article may be used for non-commercial purposes in accordance with Springer Terms and Conditions for Use of Self-Archived Versions.”

Multivariate control charts are mostly available for monitoring the process mean vector or the covariance matrix. Recently, work has been done on monitoring the multivariate coefficient of variation (CV) in Phase II of the statistical process control (SPC). However, no study has investigated the performance of the Multivariate CV charts in Phase I. The Phase I procedures are more important and involves the

estimation of the charts' limits from a historical or reference dataset that represents the in-control state of the process. In real life, contaminations are mostly present in the historical samples, hence, the Phase I procedures are mostly adopted to get rid of these contaminated samples. In this study, we investigate the performance of a variety of multivariate CV charts in Phase I considering both diffuse symmetric and localized CV disturbance scenarios, using probability to signal as a performance measure. A real-life application, concerning carbon fiber tubing, is also provided to show the implementation of the proposed charts in Phase I. The findings of this study will be useful for practitioners in their selection of an efficient Phase I control chart for monitoring multivariate CV.

4.1 Introduction

Control charts are one of the statistical process control (SPC) problem-solving tools, used to ensure process uniformity (Montgomery, 2009). The basic purpose of the implementation of control chart procedures is to detect abnormal/un-natural variations in process parameters. Control charts are mostly employed in a two-phase procedure (Jensen et al., 2006). In Phase I (retrospective phase), they are used to study a historical reference sample, which involves establishing the in-control state and evaluating the process stability to ensure that the reference sample is representative of the process (Zhou et al., 2007). Once the in-control reference sample is determined, the process parameters, if unknown, are estimated from Phase I, and control chart limits are obtained for used in Phase II. The Phase II aspect involves online monitoring of the process. If there occurs any shift in process parameters, it needs to be detected quickly so that corrective actions can be taken at an early stage.

Most of the works on control charts focus on monitoring either the process mean or the process variability, and are based on the requirements that the process mean is stable, and independent of the process standard deviation (cf. (Yeong et al., 2016)). However, in many real-life processes, the process standard

deviation is dependent on the mean, and the mean is not constant. In such operations, monitoring the process mean using the \bar{X} chart; or using S or R charts for monitoring process variability, seem to be improper. In such cases, it is more appropriate to monitor the coefficient of variation (CV).

The univariate CV is the ratio of the standard deviation to the mean of a random variable, i.e., $\gamma = \frac{\sigma}{\mu}$. As a relative dispersion measure to the mean, it is useful for comparing the variability of populations with really diverse means, and considerably straightforward to interpret in practice (Aerts et al., 2015). The use of the CV is prevalent in many applications in science, medicine, engineering, economics, etc. For example, it can be used to measure the reliability of an assay in medicine and chemistry (Reed et al., 2002). Also, in clinical trials, to account for baseline variability of measurements (Pereira et al., 2004), and in quality control, to seek production processes with minimal dispersion (George Box, 1988).

Several univariate CV control charts have been adopted in many practical applications. A primary criterion for the usage of the univariate CV chart is that the standard deviation needs to be proportional to the mean so that the CV is constant. This is usually checked by plotting the rational group CV against the mean (Kang et al., 2007). In most cases, the plot is supplemented with a regression line and is also followed by a formal test of the regression slope. If the CV is constant, the CV is independent of the mean. We refer the interested reader to Kang et al. (2007); Calzada and Scariano (2013); Zhang et al. (2014); Castagliola et al. (2013), and Yeong et al. (2017) for some work on univariate CV charts. Also, see Yeong et al. (2016) for a comprehensive review of CV charts that monitor the CV for univariate data.

When a p -dimensional normal random vector \mathbf{Y} with mean vector $\boldsymbol{\mu}_0$ and the covariance matrix ($\boldsymbol{\Sigma}_0$) is available, the calculation of the univariate CV of each random variable is not suitable, since no account is made for the correlation structure. Various multivariate extensions of the univariate CV have been proposed, all of which reduce to univariate CV when the number of variables (p) is equal to one and the mean is positive. However, as soon as p is greater than one, they do not measure the same quantity (Aerts et al., 2015). Reyment (1960) was the first to generalize the univariate theory of the relative variation to

multivariate CV. He gave a definition where the denominator denotes the squared norm of the mean, and the numerator signifies the geometric average of the eigenvalues of Σ_0 .

Van Valen (1974) examined the approach by Reyment (1960), argued that the determinant of Σ_0 will be null or close to zero if one of its eigenvalues is zero or very small. This determinant measures the area of the ellipsoids of the distribution under consideration and expresses in terms of the correlation structure of the data matrix. Voinov, V. G. and Nikulin (1996) established that a generalization of the multivariate CV can be obtained from the well-known Mahalanobis distance. They showed that this measure of variation between μ_0 and Σ_0 becomes bigger in the sense of positive definiteness as $\mu_0^T \mu_0$ becomes big. Their definition has the characteristic of being unchanged under linear transformations. Hence, changing the scale of some or all of the variables will not affect their multivariate CV. More recently, Albert and Zhang (2010) reviewed the existing multivariate CV and proposed a new multivariate CV. Their proposed scheme is well suited for high-dimensional problems, as it does not require matrix inversion, only involves the calculation of the mean vector, covariance matrix and simple quadratic forms. To check whether CV is constant for multivariate datasets, it is usually recommended to plot the square of the rational group estimated multivariate CV against $\bar{Y}^T \bar{Y}$, where \bar{Y} is the sample mean vector. This is also followed by a formal test of the regression slope (Kang et al., 2007).

Yeong et al. (2016) proposed two one-sided multivariate CV charts in Phase II. They adopted the multivariate CV definition by Voinov, V. G. and Nikulin (1996), and presented the distributional properties of the multivariate CV in Phase II, using average run length (ARL). They also employed expected average run length (EARL) when the shift size is unknown. ARL is defined as the average number of samples that is plotted on a control chart until an out-of-control signal is issued by the chart while EARL is the expected value of the ARL, integrated over the density function of the shift size. However, a major setback of the proposed two one-sided multivariate CV by Yeong et al. (2016), is that the charts are slow in detecting multivariate CV shift in the Phase-II process. To overcome this problem, Lim et al. (2017)

proposed two one-sided run sum charts for monitoring the multivariate CV in the Phase-II process. The charts are simultaneously utilized to detect positive and negative shifts in multivariate CV.

In all of the proposed multivariate CV charts, the multivariate CV statistics (γ), is given as functions of the process parameters, $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$. For example, the multivariate CV statistic by Voinov, V. G. and Nikulin (1996) given as $(\boldsymbol{\mu}_0^T \boldsymbol{\Sigma}_0^{-1} \boldsymbol{\mu}_0)^{-1/2}$, depends on the in-control process parameters. In practice, these parameters are usually unknown, and therefore, must be estimated from samples taken when the process is assumed to be in control. Specifically, the process parameters are generally estimated by the sample mean vector $\bar{\mathbf{Y}}$, and sample covariance matrix \mathbf{S} , respectively, given as follows: $\bar{\mathbf{Y}} = \frac{1}{n} \sum_{i=1}^n \mathbf{y}_i$ and $\mathbf{S} = \frac{1}{(n-1)} \sum_{i=1}^n (\mathbf{y}_i - \bar{\mathbf{y}})(\mathbf{y}_i - \bar{\mathbf{y}})^T$, where n is the sample size from the process at each time point $i = 1, 2, \dots, n$. The Phase II performance of the multivariate CV depends on the stability of the estimated parameters obtained from Phase I. An optimal performance requires any changes in these parameters to be detected as early as possible (Riaz and Schoonhoven, 2011). Thus, it is important to study the Phase I analysis of the multivariate CV chart.

This chapter aims to provide a comprehensive study of the Phase I analysis of the multivariate CV charts for monitoring multivariate CV. We investigate the Phase I performance of the charts using different multivariate CV expressions proposed in the literature. We use probability to signal (PTS) as a performance measure. The PTS performance measure has found great application in Phase I analysis of the control charts for both location and dispersion parameters. The rest of the chapter is organized as follows. In Section 4.2, we present the different multivariate CV expressions and their estimates. This section also provides the design structure of the multivariate CV charts in Phase I. We describe the simulation study in Section 4.3, and present our results and discussion in Section 4.4. In Section 4.5, we offer an illustrative example using real data and give conclusions and recommendations in Section 4.6.

4.2 Multivariate CV in Phase I

Reyment (1960), Voinov, V. G. and Nikulin (1996), and Albert and Zhang (2010), respectively, proposed different multivariate extensions of the univariate CV, which have received significant attention in the literature. Reyment (1960) proposal (hereafter denoted by γ_{RR}) is given as:

$$\gamma_{RR} = \left(\det(\mathbf{\Sigma}_0)^{1/p} / \boldsymbol{\mu}_0^T \boldsymbol{\mu}_0 \right)^{1/2} \quad (4.1)$$

The denominator of γ_{RR} represents the squared norm of the mean, given as $\boldsymbol{\mu}_0^T \boldsymbol{\mu}_0 = \sum_{i=1}^p \mu_{0i}^2$. The γ_{RR} requires positive definiteness and full rank of $\mathbf{\Sigma}_0$. Voinov, V. G. and Nikulin (1996) established multivariate CV as a generalization of the Mahalanobis distance. Their proposed multivariate CV (denoted hereafter by γ_{VN}) is given as:

$$\gamma_{VN} = \left(\boldsymbol{\mu}_0^T \mathbf{\Sigma}_0^{-1} \boldsymbol{\mu}_0 \right)^{-1/2} \quad (4.2)$$

The γ_{VN} expression is invariant under change of scale, require full rank and positive definite of $\mathbf{\Sigma}_0$. Thus, it is not-well suited for high-dimensional data because it relies on the inverse of $\mathbf{\Sigma}_0$. However, the proposed multivariate extension of the univariate CV (denoted hereafter as γ_{AZ}) by Albert and Zhang (2010) given as:

$$\gamma_{AZ} = \left(\boldsymbol{\mu}_0^T \mathbf{\Sigma}_0 \boldsymbol{\mu}_0 / (\boldsymbol{\mu}_0^T \boldsymbol{\mu}_0)^2 \right)^{1/2} \quad (4.3)$$

is well-suited for high-dimensional, as it does not require the inverse of $\mathbf{\Sigma}_0$ and always exist. Specifically, γ_{AZ} does not need the full rank of $\mathbf{\Sigma}_0$.

When the parameters are unknown, we obtain m samples each of size n observations on the quality characteristics in Phase I. Let $\mathbf{Y}_1, \mathbf{Y}_2, \mathbf{Y}_3, \dots, \mathbf{Y}_n$ be the random sample of size n from the p -variate normal distribution with parameters $\boldsymbol{\mu}_0$ and $\mathbf{\Sigma}_0$. To estimate the multivariate CV in Equations (4.1), (4.2) and (4.3), we replaced $\boldsymbol{\mu}_0$ and $\mathbf{\Sigma}_0$, by their estimated values from the m samples. Specifically, we

replaced the vector $\boldsymbol{\mu}_0$ by the sample mean vector ($\bar{\mathbf{Y}}$), and the $\boldsymbol{\Sigma}_0$ matrix is replaced by the sample covariance matrix \mathbf{S}_0 . The estimated CV from Reyment (1960); Voinov, V. G. and Nikulin (1996), and Albert and Zhang (2010) are respectively given as:

$$\hat{\gamma}_{RR} = \left(\det(\mathbf{S})^{1/p} / \bar{\mathbf{Y}}^T \bar{\mathbf{Y}} \right)^{1/2} \quad (4.4)$$

$$\hat{\gamma}_{VN} = \left(\bar{\mathbf{Y}}^T \mathbf{S}^{-1} \bar{\mathbf{Y}} \right)^{-1/2} \quad (4.5)$$

$$\hat{\gamma}_{AZ} = \left(\bar{\mathbf{Y}}^T \mathbf{S} \bar{\mathbf{Y}} / (\bar{\mathbf{Y}}^T \bar{\mathbf{Y}})^2 \right)^{1/2} \quad (4.6)$$

Hence, the control charts for the multivariate CV in Phase I are time-ordered plots of the estimated CV from a given number of historical samples.

To obtain the one-sided upper control limit used in this study, we define a statistic $V_j = \frac{\hat{\gamma}_j}{\gamma}$; $\forall j = RR, VN, AZ$, where $\hat{\gamma}_j$; $\forall j = RR, VN, AZ$ is the estimated coefficient of variation defined above, and γ is the process coefficient of variation value. By taking the expectation V_j on both sides, we get $E(V_j) = \frac{E(\hat{\gamma}_j)}{\gamma} = d_{2,j,m,n}$. Where, for a specific multivariate CV estimator, d_2 entirely depends on sample size (n), and the number of Phase I samples (m). The $E(\hat{\gamma}_j)$ can be replaced with the average of sample $\hat{\gamma}_j$'s (given as $\bar{\hat{\gamma}}_j = \frac{\sum_{i=1}^m \hat{\gamma}_{j,i}}{m}$) computed from the Phase I samples. Hence, an unbiased estimator of multivariate CV can be defined as $\hat{\gamma}_{j,n} = \frac{\bar{\hat{\gamma}}_j}{d_{2,j,n}}$; $\forall j = RR, VN, AZ$.

Using these notations, we set the one-sided upper probability limits for the Phase I multivariate CV charts, considered in this study as $\widehat{UCL}_{j,m,n} = V_{m,n,j} \frac{\bar{\hat{\gamma}}_j}{d_{2,j,m,n}}$, where $V_{m,n,j}$ is a control chart constant, used to fix the false alarm probability (α) at a particular level. Let W_i be the event that $\hat{\gamma}_i$ is above \widehat{UCL} . Hence, $P(W_i|\hat{\gamma})$ is the probability that sample i generates a signal given $\hat{\gamma}$, i.e., $P(\hat{\gamma}_i > \widehat{UCL}|\hat{\gamma}) = \alpha^*$;

where α^* is the probability to signal for a single sample. The overall false alarm probability (FAP) is calculated as $\alpha = 1 - (1 - \alpha^*)^m$. For the rest of this chapter, control chart structures based on the multivariate CV estimators $\hat{\gamma}_{RR}$, $\hat{\gamma}_{VN}$, and $\hat{\gamma}_{AZ}$ will be referred as MCV_R , MCV_{VN} , and MCV_{AZ} charts, respectively.

4.3 Performance evaluation in Phase I

For the performance evaluation of control charts, the two well-known measures are average run length (ARL) and probability to signal (PTS). ARL is well suited for Phase II control charts while for Phase I charts (Adegoke et al., 2018b), one is mostly interested in the probability of detecting inconsistent/contaminated samples (or observations). Hence, in this study, PTS is used to compare the performance of the three multivariate CV control charts in Phase I. At first, we need to simulate the in-control process for obtaining control chart constants $d_{2,m,n}$ and $V_{m,n}$ to achieve the desired $FAP(\alpha)$ using the control limits for Phase I reference samples. For finding these constants at different levels of m and n , the following steps are taken:

- 100,000 Phase I datasets comprising m samples of size n are generated from the in-control multivariate normal process, i.e., $Np(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0(\gamma))$.
- The sample mean vector $\bar{\mathbf{Y}}$ and the sample covariance matrix \mathbf{S} are estimated for the m samples in each set.
- Using these $\bar{\mathbf{Y}}$ and \mathbf{S} for the m samples, the multivariate CV is estimated by all the three estimators (defined in Equations (4.4-4.6)).
- $d_{2,m,n}$ values are obtained as the mean of these 100,000 estimated multivariate CV's and are provided in Table 4.1 for $m = 30$ and 75 using $n = 5, 10$ and 15 .

- Moreover, for each set, the estimated CV's are plotted against the upper control limit (defined in Section 4.2).
- The control limit constant $V_{m,n}$ is chosen to set the desired $FAP(\alpha)$. The $V_{m,n}$ choices for $m = 30$ and 75 using $n = 5, 10$ and 15 are also provided in Table 4.1, considering $\alpha = 0.01$.

After fixing the control limits for specific choices of m and n , the performance of the three charts is evaluated using probability to signal in the presence of localized CV and diffuse symmetric CV disturbances, following Riaz and Schoonhoven (2011). These contaminated scenarios are briefly described below:

- A model for localized CV disturbances in which m_0 (out of m) samples are drawn from in-control process $Np(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0(\gamma))$, and the remaining $m_1 = m - m_0$ samples are drawn from out-of-control process $Np(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0(\delta\gamma))$.
- A model for diffuse symmetric CV disturbances in which each observation in m subgroups of size n has $a\%$ probability of being drawn from $Np(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0(\gamma))$, and $b = 100 - a\%$ probability of being drawn from $Np(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0(\delta\gamma))$.

Here, δ refers to the size of the shift in the in-control multivariate CV. The probability to signal is computed for both these scenarios considering a variety of shift levels. The steps taken for the computation of PTS are given below:

- 100,000 Phase I datasets comprising m samples of size n are simulated from the contaminated scenarios at each level of shift δ .
- The multivariate CV estimates were computed for each dataset using Equations (4.4-4.6) and plotted against the upper control limit using the control chart constants $d_{2,m,n}$ and $V_{m,n}$ provided in Table 4.1

Table 4.1: The $d_2(n)$ and V_n values for different values of m and n for all the three charts when $\alpha = 0.01$

n		$m = 30$			$m = 75$		
		VN	AZ	R	VN	AZ	R
5	V_n	2.175	2.295	1.776	2.297	2.425	1.866
	$d_2(n)$	0.79846	0.94060	0.83156	0.79892	0.94105	0.83169
10	V_n	1.793	1.849	1.551	1.864	1.921	1.598
	$d_2(n)$	0.91455	0.97329	0.92879	0.91423	0.97301	0.92851
15	V_n	1.626	1.664	1.443	1.696	1.729	1.481
	$d_2(n)$	0.94532	0.98234	0.95438	0.94549	0.98249	0.95463

- The PTS is then computed as the proportion of sample CV's plotted above the upper control limit.

A chart with higher probability to detect inconsistencies/contaminations in the Phase I samples will be treated better than others. For the localized CV disturbances, the probability of signaling values are calculated using $m_1 = 3, 6, 9$ and 12 , when $m = 30$, and $m_1 = 7, 15, 22$ and 30 , when $m = 75$. For the diffuse symmetric CV disturbances, the PTS is calculated using $a = 5, 10, 15$ and 20 , for $m = 30$ and 75 . These PTS results are provided in Tables 4.2 - 4.3 and Tables 4.4 - 4.5, for the localized CV and diffuse symmetric CV disturbances, respectively. Moreover, for better visual comparison, the PTS of the three multivariate CV charts are compared in Figures 4.1 - 4.3. In each graph, the PTS of each multivariate CV chart is plotted against shift δ . Figures 4.1 - 4.2 presents the comparison using $n = 10$, $m = 30$ at different levels of m_1 and a in presence of localized CV and diffuse symmetric CV disturbances, respectively. The effect of sample size on the performance of the MCVVN chart, is presented in Figure 4.3 using $m = 30$ at different levels of n , in presence of localized CV (using $m_1 = 6$) and diffuse symmetric CV ($b = 10\%$) disturbances. Similar comparisons can be made for the MCV_R and MCV_{VZ} charts.

4.4 Results and discussion

In this section, we provide a comprehensive study of the Phase I analysis of the multivariate CV control charts considered in this study. For a fixed FAP $\alpha = 0.01$, the chart that yields the highest probability

of signaling is considered better than the other control charts. We aim to recommend the choice of the multivariate CV control chart that gives the best Phase I performance. The effects of the shift size, the sample size (n) and m_1 on the proposed charts are also studied. The significant findings of the Phase I analysis are summarized below:

- 1 In the presence of localized CV disturbances, the MCV_R chart, followed by the MCV_{VN} chart, dominantly performed better than the MCV_{AZ} chart, especially for small to moderate shift in the in-control CV. This is true for all values of n , m , and m_1 . For example, when $n = 10$, $\delta = 1.6$, and $m_1 = 6$, in Table 4.2, the PTS values are 0.3629, 0.4048 and 0.6440, for the MCV_{AZ} , MCV_{VN} and MCV_R charts, respectively. Figure 4.1 provides a graphical comparison of the three charts in presence of localized CV disturbances for $n = 10$ and $m = 30$, when $m_1 = 6$ and 12, respectively.
- 2 In the presence of diffuse symmetric CV disturbances, the MCV_{VN} chart, followed very closely by the MCV_R , chart, dominantly performed better than the MCV_{AZ} chart, also for small to moderate shift in the in-control CV value and small values of b . For example, when $n = 10$, $\delta = 1.6$, and $b = 10$, in Table 4.4, the PTS values are 0.0014, 0.0572 and 0.0460, for the MCV_{AZ} , MCV_{VN} and MCV_R charts, respectively. Figure 4.2 provides a graphical comparison of the three multivariate CV charts in presence of diffused symmetric CV disturbances for $n = 10$ and $m = 30$, for $b(\%) = 10$ and 15, respectively.
- 3 Clearly observed from Figures 4.1 - 4.2 the probability of signaling for all the multivariate CV charts approaches one, as the size of the shifts increases in the presence of both localized CV and diffuse symmetric CV disturbances.
- 4 It can be observed from Figure 4.3 that for any given value of m_1 or $b\%$, greater PTS is observed when large values of n are used (also see Tables 4.2- 4.5).
- 5 Increasing the size of m increases the PTS of all the multivariate CV charts.

Table 4.2: The PTS values of the localized CV disturbances model for $m = 30$, $\alpha = 0.01$ at different levels of n and m_1

		<i>VN</i>				<i>AZ</i>				<i>R</i>			
<i>n</i>		m_1				m_1				m_1			
5	δ	3	6	9	12	3	6	9	12	3	6	9	12
	1.0	0.0103	0.0100	0.0100	0.0098	0.0095	0.0096	0.0103	0.0102	0.0098	0.0100	0.0096	0.0096
	1.2	0.0162	0.0207	0.0254	0.0258	0.0154	0.0183	0.0199	0.0210	0.0229	0.0298	0.0302	0.0311
	1.4	0.0599	0.0786	0.0781	0.0722	0.0430	0.0565	0.0587	0.0519	0.0915	0.1148	0.1128	0.1025
	1.6	0.1490	0.1853	0.1784	0.1513	0.1084	0.1337	0.1271	0.1122	0.2321	0.2798	0.2670	0.2192
	1.8	0.2680	0.3267	0.2997	0.2501	0.1961	0.2404	0.2237	0.1849	0.4002	0.4761	0.4482	0.3627
	2.0	0.3950	0.4676	0.4313	0.3570	0.3001	0.3566	0.3294	0.2689	0.5540	0.6439	0.6073	0.5061
	2.5	0.6461	0.7362	0.6995	0.5955	0.5361	0.6150	0.5697	0.4682	0.7858	0.8762	0.8548	0.7671
	3.0	0.7806	0.8669	0.8447	0.7513	0.6881	0.7752	0.7352	0.6273	0.8720	0.9502	0.9444	0.8909
	4.0	0.8829	0.9529	0.9456	0.8966	0.8273	0.9056	0.8863	0.8070	0.9276	0.9841	0.9859	0.9687
	5.0	0.9149	0.9746	0.9733	0.9448	0.8792	0.9462	0.9375	0.8832	0.9422	0.9912	0.9936	0.9859
10	1.0	0.0102	0.0096	0.0103	0.0103	0.0104	0.0097	0.0104	0.0105	0.0103	0.0100	0.0097	0.0096
	1.2	0.0266	0.0346	0.0395	0.0358	0.0246	0.0327	0.0346	0.0334	0.0429	0.0574	0.0567	0.0516
	1.4	0.1323	0.1685	0.1623	0.1340	0.1148	0.1462	0.1432	0.1201	0.2460	0.2955	0.2756	0.2283
	1.6	0.3390	0.4048	0.3747	0.3088	0.2954	0.3629	0.3328	0.2686	0.5602	0.6440	0.6037	0.4994
	1.8	0.5527	0.6444	0.6033	0.4973	0.5044	0.5874	0.5441	0.4495	0.7707	0.8644	0.8400	0.7432
	2.0	0.7105	0.8063	0.7735	0.6642	0.6663	0.7601	0.7188	0.6104	0.8726	0.9505	0.9448	0.8871
	2.5	0.8821	0.9573	0.9522	0.8987	0.8599	0.9397	0.9290	0.8618	0.9414	0.9926	0.9954	0.9877
	3.0	0.9284	0.9859	0.9878	0.9685	0.9170	0.9796	0.9802	0.9510	0.9534	0.9970	0.9992	0.9981
	4.0	0.9503	0.9956	0.9980	0.9949	0.9465	0.9940	0.9965	0.9908	0.9571	0.9984	0.9998	0.9998
	5.0	0.9547	0.9972	0.9992	0.9983	0.9531	0.9965	0.9987	0.9968	0.9574	0.9986	0.9999	0.9999
15	1.0	0.0104	0.0097	0.0102	0.0097	0.0099	0.0101	0.0097	0.0101	0.0098	0.0098	0.0099	0.0101
	1.2	0.0436	0.0588	0.0617	0.0581	0.0416	0.0585	0.0630	0.0573	0.0730	0.0993	0.0972	0.0844
	1.4	0.2344	0.2898	0.2793	0.2301	0.2165	0.2728	0.2622	0.2194	0.4358	0.5106	0.4738	0.3811
	1.6	0.5301	0.6221	0.5855	0.4890	0.5028	0.5930	0.5583	0.4635	0.7733	0.8700	0.8495	0.7540
	1.8	0.7490	0.8490	0.8226	0.7241	0.7227	0.8216	0.7969	0.6944	0.9004	0.9717	0.9718	0.9370
	2.0	0.8559	0.9417	0.9339	0.8708	0.8394	0.9282	0.9175	0.8465	0.9383	0.9917	0.9947	0.9862
	2.5	0.9369	0.9908	0.9939	0.9839	0.9320	0.9885	0.9914	0.9777	0.9559	0.9980	0.9997	0.9996
	3.0	0.9520	0.9965	0.9989	0.9974	0.9502	0.9959	0.9984	0.9960	0.9574	0.9986	0.9999	1.0000
	4.0	0.9569	0.9983	0.9998	0.9998	0.9564	0.9981	0.9997	0.9996	0.9576	0.9987	1.0000	1.0000
	5.0	0.9574	0.9986	0.9999	0.9999	0.9572	0.9985	0.9999	0.9999	0.9576	0.9988	1.0000	1.0000

Table 4.3: The PTS values of the localized CV disturbances model for $m = 75$, $\alpha = 0.01$ at different levels of n and m_1

		VN				AZ				R			
n		m_1				m_1				m_1			
5	δ	7	15	22	30	7	15	22	30	7	15	22	30
	1.0	0.0101	0.0103	0.0105	0.0096	0.0101	0.0096	0.0101	0.0104	0.0096	0.0098	0.01	0.01
	1.2	0.0230	0.0265	0.0282	0.0290	0.0215	0.0239	0.0264	0.0244	0.0275	0.0366	0.0402	0.0387
	1.4	0.0883	0.1160	0.1128	0.0968	0.0697	0.0905	0.0864	0.0738	0.1398	0.1771	0.1699	0.1435
	1.6	0.2360	0.2937	0.2716	0.2220	0.1752	0.2242	0.2090	0.1717	0.3593	0.4317	0.4020	0.3232
	1.8	0.4292	0.5058	0.4626	0.3786	0.3307	0.3972	0.3635	0.2948	0.6092	0.6896	0.6458	0.5290
	2.0	0.6147	0.6963	0.6479	0.5324	0.4976	0.5705	0.5277	0.4277	0.7890	0.8607	0.8237	0.7091
	2.5	0.8792	0.9332	0.9067	0.8118	0.7917	0.8605	0.8177	0.6972	0.9586	0.9865	0.9788	0.9353
	3.0	0.9584	0.9859	0.9773	0.9321	0.9136	0.9574	0.9355	0.8539	0.9880	0.9984	0.9975	0.9872
	4.0	0.9904	0.9987	0.9980	0.9896	0.9776	0.9942	0.9902	0.9650	0.9971	0.9999	0.9999	0.9992
	5.0	0.9956	0.9997	0.9996	0.9975	0.9901	0.9985	0.9975	0.9887	0.9984	1.0000	1.0000	0.9999
10	1.0	0.0098	0.0101	0.0104	0.0102	0.0101	0.0104	0.0097	0.0099	0.01	0.0103	0.01	0.0099
	1.2	0.0401	0.0535	0.0551	0.0526	0.0351	0.0486	0.0494	0.0483	0.0620	0.0843	0.0849	0.0748
	1.4	0.2127	0.2679	0.2430	0.2009	0.1846	0.2389	0.2209	0.1812	0.3919	0.4664	0.4311	0.3427
	1.6	0.5305	0.6182	0.5740	0.4595	0.4804	0.5615	0.5190	0.4105	0.7899	0.8662	0.8290	0.7085
	1.8	0.7933	0.8667	0.8271	0.7087	0.7437	0.8217	0.7767	0.6527	0.9514	0.9837	0.9752	0.9233
	2.0	0.9196	0.9652	0.9469	0.8696	0.8905	0.9448	0.9192	0.8271	0.9877	0.9984	0.9976	0.9860
	2.5	0.9900	0.9989	0.9983	0.9892	0.9851	0.9977	0.9962	0.9797	0.9983	1.0000	1.0000	0.9999
	3.0	0.9972	0.9999	0.9999	0.9992	0.9959	0.9998	0.9998	0.9979	0.9991	1.0000	1.0000	1.0000
	4.0	0.9989	1.0000	1.0000	1.0000	0.9987	1.0000	1.0000	1.0000	0.9993	1.0000	1.0000	1.0000
	5.0	0.9992	1.0000	1.0000	1.0000	0.9991	1.0000	1.0000	1.0000	0.9993	1.0000	1.0000	1.0000
15	1.0	0.0099	0.0095	0.0103	0.0096	0.0103	0.0102	0.0102	0.0099	0.0097	0.0102	0.01	0.0103
	1.2	0.0578	0.0804	0.0826	0.0719	0.0527	0.0739	0.0754	0.0660	0.1105	0.1459	0.1403	0.1170
	1.4	0.3508	0.4326	0.4062	0.3216	0.3160	0.3935	0.3696	0.2898	0.6420	0.7281	0.6828	0.5570
	1.6	0.7456	0.8302	0.7923	0.6698	0.7090	0.7920	0.7470	0.6239	0.9505	0.9842	0.9754	0.9219
	1.8	0.9349	0.9745	0.9613	0.8946	0.9165	0.9626	0.9440	0.8619	0.9930	0.9995	0.9994	0.9951
	2.0	0.9825	0.9971	0.9953	0.9757	0.9765	0.9949	0.9916	0.9627	0.9980	1.0000	1.0000	0.9998
	2.5	0.9978	1.0000	1.0000	0.9997	0.9973	0.9999	0.9999	0.9993	0.9993	1.0000	1.0000	1.0000
	3.0	0.9990	1.0000	1.0000	1.0000	0.9989	1.0000	1.0000	1.0000	0.9993	1.0000	1.0000	1.0000
	4.0	0.9993	1.0000	1.0000	1.0000	0.9993	1.0000	1.0000	1.0000	0.9994	1.0000	1.0000	1.0000
	5.0	0.9993	1.0000	1.0000	1.0000	0.9993	1.0000	1.0000	1.0000	0.9994	1.0000	1.0000	1.0000

4.5 Illustrative example

In this section, we provide an illustrative example to show the applications of the proposed multivariate CV charts in real life situation, using the carbon fiber tubing dataset. The carbon fiber tubes are mostly composed of carbon atoms and have numerous properties which include high stiffness, high tensile strength, high chemical resistance, high-temperature tolerance, low weight, and low thermal expansion (Zhang et al., 2016). These features have made carbon fiber prevalent in aerospace, beam constructions,

Table 4.4: The PTS values of the diffused symmetric CV disturbances model for $m = 30$, $\alpha = 0.01$ at different levels of n and b

		VN				AZ				R			
n		$b\%$				$b\%$				$b\%$			
5	δ	5	10	15	20	5	10	15	20	5	10	15	20
	1.0	0.0098	0.0097	0.0097	0.0104	0.0100	0.0101	0.0103	0.0099	0.0099	0.0099	0.0102	0.0095
	1.2	0.0101	0.0104	0.0114	0.0117	0.0000	0.0000	0.0000	0.0000	0.0139	0.0142	0.0146	0.0155
	1.4	0.0167	0.0221	0.0240	0.0269	0.0001	0.0002	0.0002	0.0001	0.0178	0.0220	0.0265	0.0307
	1.6	0.0350	0.0493	0.0548	0.0579	0.0002	0.0006	0.0011	0.0015	0.0288	0.0410	0.0507	0.0583
	1.8	0.0616	0.0912	0.1036	0.1060	0.0012	0.0025	0.0025	0.0035	0.0445	0.0694	0.0838	0.0993
	2.0	0.1046	0.1544	0.1665	0.1679	0.0028	0.0062	0.0087	0.0090	0.0654	0.1095	0.1341	0.1498
	2.5	0.2425	0.3302	0.3468	0.3370	0.0150	0.0327	0.0407	0.0426	0.1503	0.2396	0.2825	0.3062
	3.0	0.3962	0.5036	0.5084	0.4849	0.0399	0.0787	0.1005	0.1073	0.2547	0.3798	0.4338	0.4624
	4.0	0.6347	0.7259	0.7196	0.6751	0.1289	0.2273	0.2645	0.2719	0.4528	0.6029	0.6580	0.6837
	5.0	0.7680	0.8369	0.8204	0.7760	0.2422	0.3794	0.4267	0.4335	0.6029	0.7343	0.7809	0.8020
10	1.0	0.0103	0.0100	0.0098	0.0103	0.0097	0.0099	0.0103	0.0101	0.0103	0.0104	0.0103	0.0098
	1.2	0.0108	0.0112	0.0130	0.0142	0.0000	0.0000	0.0000	0.0000	0.0101	0.0113	0.0116	0.0125
	1.4	0.0177	0.0265	0.0298	0.0335	0.0000	0.0003	0.0002	0.0002	0.0161	0.0221	0.0254	0.0275
	1.6	0.0381	0.0572	0.0642	0.0707	0.0005	0.0014	0.0014	0.0020	0.0301	0.0460	0.0538	0.0598
	1.8	0.0776	0.1136	0.1260	0.1264	0.0028	0.0056	0.0066	0.0067	0.0595	0.0885	0.1094	0.1170
	2.0	0.1380	0.1903	0.2036	0.2020	0.0065	0.0143	0.0163	0.0167	0.1031	0.1556	0.1821	0.1956
	2.5	0.3420	0.4210	0.4304	0.3984	0.0430	0.0720	0.0816	0.0768	0.2599	0.3727	0.4165	0.4270
	3.0	0.5488	0.6197	0.6092	0.5648	0.1229	0.1904	0.2031	0.1821	0.4365	0.5712	0.6185	0.6196
	4.0	0.7999	0.8417	0.8108	0.7569	0.3415	0.4637	0.4751	0.4323	0.7039	0.8118	0.8389	0.8304
	5.0	0.9063	0.9229	0.8937	0.8430	0.5362	0.6670	0.6717	0.6192	0.8404	0.9075	0.9248	0.9151
15	1.0	0.0102	0.0102	0.0100	0.0103	0.0100	0.0099	0.0103	0.0097	0.0097	0.0097	0.0096	0.0101
	1.2	0.0134	0.0149	0.0156	0.0168	0.0001	0.0000	0.0001	0.0001	0.0124	0.0141	0.0138	0.0147
	1.4	0.0250	0.0326	0.0369	0.0407	0.0003	0.0003	0.0004	0.0007	0.0198	0.0276	0.0303	0.0334
	1.6	0.0486	0.0694	0.0808	0.0855	0.0010	0.0018	0.0020	0.0027	0.0362	0.0610	0.0709	0.0765
	1.8	0.0948	0.1335	0.1492	0.1490	0.0051	0.0080	0.0081	0.0101	0.0702	0.1163	0.1346	0.1475
	2.0	0.1611	0.2233	0.2349	0.2283	0.0132	0.0225	0.0246	0.0238	0.1209	0.1938	0.2198	0.2332
	2.5	0.3921	0.4768	0.4766	0.4434	0.0650	0.1104	0.1160	0.1064	0.3228	0.4455	0.4782	0.4773
	3.0	0.6124	0.6815	0.6613	0.6084	0.1754	0.2575	0.2620	0.2313	0.5296	0.6624	0.6922	0.6710
	4.0	0.8594	0.8815	0.8492	0.7907	0.4659	0.5801	0.5607	0.4889	0.8042	0.8821	0.8895	0.8688
	5.0	0.9437	0.9485	0.9195	0.8668	0.6882	0.7771	0.7501	0.6634	0.9124	0.9532	0.9536	0.9379

and motorsports, among others.

Pultrusion or filament winding process is commonly used to manufacture the carbon fiber tube. The pultrusion is a continuous process of fabricating composite materials. Pultruded tubes are the most cost-effective approach to making continuous hollow tubes, which can be achieved by following a sequence of operations. The first step in the operations is continuous fiberglass reinforcement, which involves rolling of the filament or fabric to maintain strength across the profile. The fabrics are then attached to a machine known as a tension roller, which will help to shape it as intended. The process is then introduced into

Table 4.5: The PTS values of the diffused symmetric CV disturbances model for $m = 75$, $\alpha = 0.01$ at different levels of n and b

		VN				AZ				R			
n		$b\%$				$b\%$				$b\%$			
5	δ	5	10	15	20	5	10	15	20	5	10	15	20
	1.0	0.0104	0.0100	0.0100	0.0102	0.0101	0.0099	0.0098	0.0104	0.0100	0.0103	0.0100	0.0103
	1.2	0.0118	0.0142	0.0152	0.0162	0.0001	0.0001	0.0001	0.0001	0.0132	0.0138	0.0145	0.0160
	1.4	0.0244	0.0316	0.0404	0.0430	0.0002	0.0001	0.0001	0.0003	0.0196	0.0256	0.0312	0.0359
	1.6	0.0545	0.0791	0.0968	0.0990	0.0004	0.0008	0.0010	0.0015	0.0388	0.0556	0.0697	0.0813
	1.8	0.1063	0.1613	0.1829	0.1822	0.0016	0.0038	0.0051	0.0054	0.0677	0.1057	0.1340	0.1482
	2.0	0.1922	0.2628	0.2924	0.2848	0.0047	0.0087	0.0116	0.0145	0.1076	0.1754	0.2200	0.2432
	2.5	0.4416	0.5481	0.5656	0.5447	0.0311	0.0552	0.0692	0.0740	0.2666	0.4015	0.4715	0.5030
	3.0	0.6613	0.7593	0.7602	0.7303	0.0846	0.1552	0.1857	0.1879	0.4479	0.6124	0.6811	0.7105
	4.0	0.8909	0.9361	0.9263	0.9018	0.2666	0.4186	0.4763	0.4793	0.7243	0.8530	0.8970	0.9125
	5.0	0.9625	0.9799	0.9725	0.9560	0.4650	0.6384	0.6992	0.7009	0.8646	0.9412	0.9630	0.9696
10	1.0	0.0103	0.0096	0.0096	0.0104	0.0101	0.0104	0.0101	0.0100	0.0102	0.0102	0.0097	0.0104
	1.2	0.0136	0.0162	0.0173	0.0183	0.0000	0.0000	0.0000	0.0000	0.0121	0.0145	0.0156	0.0182
	1.4	0.0284	0.0381	0.0448	0.0499	0.0001	0.0002	0.0004	0.0003	0.0214	0.0309	0.0373	0.0438
	1.6	0.0634	0.0949	0.1047	0.1136	0.0007	0.0017	0.0021	0.0026	0.0453	0.0748	0.0926	0.1081
	1.8	0.1352	0.1920	0.2094	0.2126	0.0028	0.0072	0.0088	0.0102	0.0984	0.1580	0.1921	0.2079
	2.0	0.2461	0.3230	0.3395	0.3362	0.0114	0.0226	0.0283	0.0272	0.1749	0.2739	0.3230	0.3435
	2.5	0.5787	0.6688	0.6681	0.6252	0.0757	0.1307	0.1453	0.1359	0.4429	0.6008	0.6606	0.6700
	3.0	0.8136	0.8692	0.8554	0.8117	0.2145	0.3298	0.3516	0.3213	0.6908	0.8258	0.8634	0.8637
	4.0	0.9717	0.9813	0.9706	0.9471	0.5720	0.7153	0.7269	0.6749	0.9257	0.9731	0.9807	0.9781
	5.0	0.9950	0.9965	0.9921	0.9804	0.8012	0.9028	0.9011	0.8607	0.9818	0.9950	0.9968	0.9955
15	1.0	0.0100	0.0097	0.0098	0.0102	0.0101	0.0104	0.0101	0.0101	0.0097	0.0098	0.0103	0.0098
	1.2	0.0147	0.0162	0.0180	0.0200	0.0000	0.0000	0.0000	0.0000	0.0121	0.0137	0.0134	0.0162
	1.4	0.0286	0.0401	0.0456	0.0486	0.0001	0.0002	0.0001	0.0004	0.0213	0.0337	0.0388	0.0450
	1.6	0.0656	0.0967	0.1121	0.1187	0.0010	0.0013	0.0020	0.0024	0.0532	0.0842	0.1011	0.1162
	1.8	0.1436	0.1967	0.2156	0.2218	0.0043	0.0081	0.0091	0.0104	0.1134	0.1797	0.2185	0.2336
	2.0	0.2624	0.3375	0.3549	0.3477	0.0145	0.0272	0.0287	0.0312	0.2104	0.3129	0.3722	0.3827
	2.5	0.6180	0.6992	0.6874	0.6498	0.0988	0.1578	0.1639	0.1523	0.5388	0.6863	0.7296	0.7245
	3.0	0.8626	0.8934	0.8734	0.8281	0.2904	0.3989	0.3985	0.3472	0.7980	0.8959	0.9128	0.9008
	4.0	0.9860	0.9885	0.9780	0.9542	0.7132	0.8074	0.7838	0.6978	0.9714	0.9907	0.9915	0.9872
	5.0	0.9984	0.9983	0.9945	0.9836	0.9083	0.9518	0.9324	0.8739	0.9954	0.9989	0.9989	0.9977

a resin mixture, which will soak and permeate it, and the resin base is exposed to a heat source. After it, the profile is moved along a pull mechanism to meet the cutting saw, where it is cut into appropriate lengths. This completes the pultrusion process. The pultrusion process and filament winding approach, although are the most common, but are not the only way to manufacture the tubes. Some other methods like bladder molding, vacuum infusion, compression molding, and autoclave processing can also be used to manufacture carbon fiber tubes.

The data used for illustration was adopted by Santos-Fernandez (2012); Santos-Fernandez, Edgar and

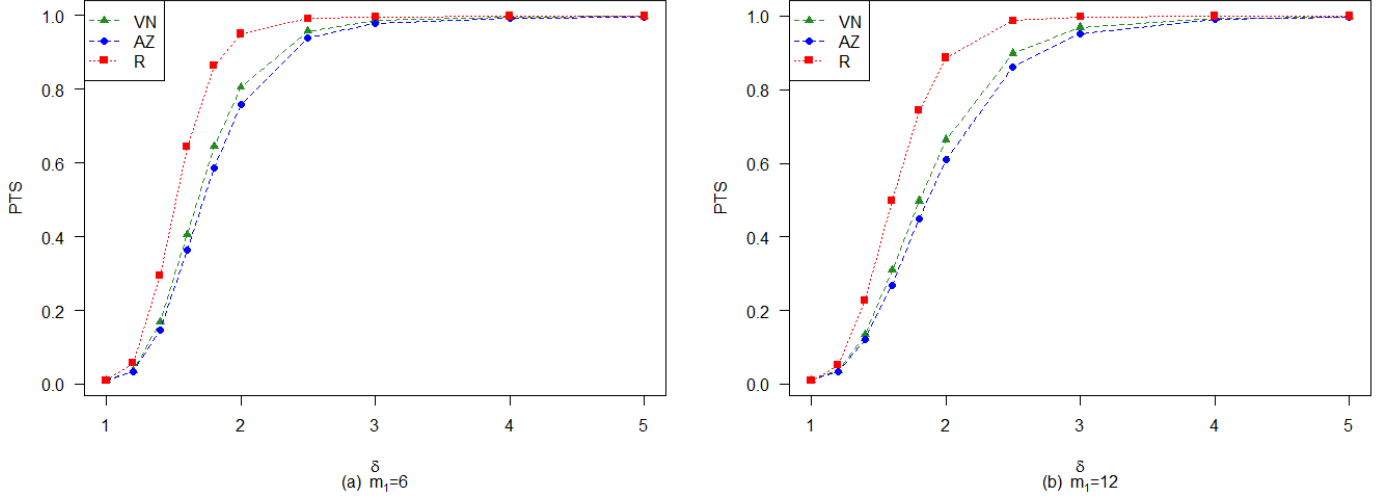


Figure 4.1: PTS comparison of the Multivariate CV charts in presence of localized CV disturbances when $n = 10$, $m = 30$ and $\alpha = 0.01$ at different levels of m_1 .

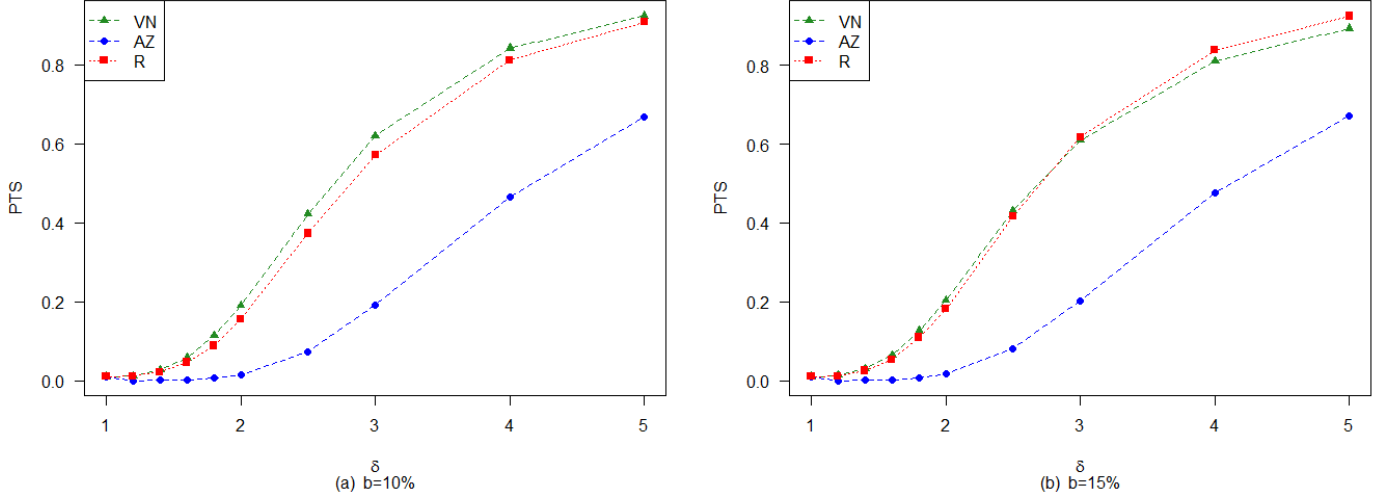


Figure 4.2: PTS comparison of Multivariate CV charts in presence of diffuse CV disturbances when $n = 10$, $m = 30$ and $\alpha = 0.01$ at different levels of b .

Santos-Fernandez (2016), and comprises three variables, among which we have used the inner diameter of the tubes (\mathbf{Y}_1) and the tubes' length (\mathbf{Y}_2), which are measured in a specific carbon fiber tube. The

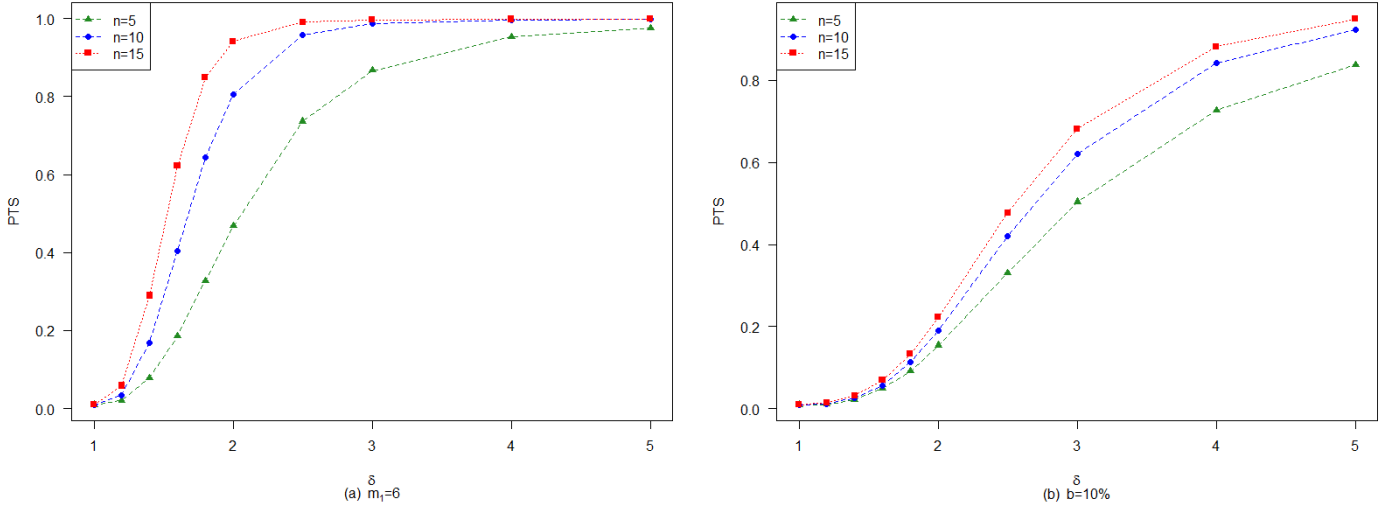


Figure 4.3: PTS comparison for the Multivariate VN chart when $m = 30$ and $a\alpha = 0.01$ at different levels of n .

mean vectors and covariance matrix of each sample are represented as follows:

$$\boldsymbol{\mu}_0 = \begin{pmatrix} \boldsymbol{\mu}_{1t} \\ \boldsymbol{\mu}_{2t} \end{pmatrix}, \quad \text{and} \quad \boldsymbol{\Sigma}_0 = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \text{ where } 1, 2, \dots, 30,$$

where $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$ are the population mean vectors, σ_1^2 and σ_2^2 are the population variance for (\mathbf{Y}_1) and (\mathbf{Y}_2) , respectively. These parameters are usually unknown (as is the case in our example) and needs to be estimated from sample information. In this study, we used 30 Phase I samples of size $n = 5$. The estimated mean vectors ($\bar{\mathbf{Y}}_{1t}$, and $\bar{\mathbf{Y}}_{2t}$), the statistics $\bar{\mathbf{Y}}^T \bar{\mathbf{Y}}$, the estimated variances-covariances (\mathbf{S}_1 and \mathbf{S}_2), and multivariate CV estimates ($\hat{\gamma}_{RR}$, $\hat{\gamma}_{VN}$, and $\hat{\gamma}_{AZ}$) for each sample are provided in Table 4.6. The subscript t is used to denote a given time point (or the sample number).

The first step for the implementation of CV charts is to check the constancy of multivariate CV for the given Phase I dataset. For this, it is usually recommended to plot $\bar{\mathbf{Y}}^T \bar{\mathbf{Y}}$ against $\hat{\gamma}^2$ (Lim et al., 2017). Figure 4.4 shows the scatter plots of the $\bar{\mathbf{Y}}^T \bar{\mathbf{Y}}$ against the estimated CV-squared for all the three multivariate CV charts based on the statistics defined in Equations (4.4) - (4.6). We observe from Figure

Table 4.6: Sample information and CV estimators for the carbon fiber tubing data

<i>SampleNo.</i>	$\bar{\mathbf{Y}}_1$	$\bar{\mathbf{Y}}_2$	$\bar{\mathbf{Y}}^T \bar{\mathbf{Y}}$	\mathbf{S}_1^2	\mathbf{S}_2^2	\mathbf{S}_{12}	$\hat{\gamma}_{VN}$	$\hat{\gamma}_{AZ}$	$\hat{\gamma}_{RR}$
1	1.026	50.152	2516.276	0.00153	0.02642	0.00201	0.003149	0.003245	0.001548
2	0.976	49.96	2496.954	0.00663	0.0714	0.01395	0.004275	0.005367	0.002586
3	1.024	50.172	2518.278	0.00103	0.02672	0.00369	0.002491	0.003266	0.001217
4	0.998	49.978	2498.796	0.00337	0.03037	-0.00676	0.002494	0.00347	0.001736
5	0.96	49.788	2479.767	0.0049	0.03772	0.01275	0.001425	0.003925	0.001379
6	1.004	50.012	2502.208	0.00233	0.01657	0.00104	0.002557	0.002576	0.001565
7	0.97	50.036	2504.542	0.00195	0.02963	-0.00047	0.003408	0.003438	0.00174
8	1.02	50.12	2513.055	0.00185	0.06355	-0.00723	0.003463	0.005016	0.001794
9	0.992	50.032	2504.185	0.00227	0.02542	-0.00268	0.002908	0.003179	0.001685
10	0.994	50.13	2514.005	0.00218	0.0576	0.0003	0.004774	0.004786	0.002111
11	1.004	49.972	2498.209	0.00203	0.05387	0.002615	0.004592	0.004647	0.002013
12	1.016	49.97	2498.033	0.00153	0.0204	0.00305	0.002491	0.002866	0.001369
13	0.978	50.086	2509.564	0.00207	0.11663	0.012215	0.004738	0.00683	0.001956
14	0.99	49.96	2496.982	0.00035	0.07615	0.000325	0.005383	0.005522	0.001436
15	1.008	50.062	2507.22	0.00212	0.04632	0.009105	0.001856	0.004314	0.001249
16	0.968	49.942	2495.14	0.00177	0.03277	0.005905	0.002444	0.003636	0.001388
17	1.002	49.998	2500.804	0.00027	0.10237	0.001255	0.006323	0.006398	0.001429
18	1.01	50.036	2504.621	0.00085	0.05338	0.006275	0.001968	0.004627	0.000989
19	1.014	49.986	2499.628	0.00023	0.07333	-0.00106	0.004562	0.005414	0.00126
20	1.008	50.024	2503.417	0.00057	0.00793	-0.00097	0.001531	0.001775	0.00087
21	0.962	49.846	2485.549	0.00307	0.06803	0.01406	0.001327	0.005252	0.00116
22	0.98	50.162	2517.187	0.00275	0.10812	0.01605	0.002701	0.006572	0.001582
23	0.934	49.828	2483.702	0.00253	0.00717	0.00096	0.001667	0.001703	0.001293
24	0.992	50.092	2510.193	0.00237	0.04467	0.007645	0.003011	0.004232	0.001656
25	0.988	49.956	2496.578	0.00377	0.04123	0.00664	0.003558	0.004076	0.002056
26	0.978	49.896	2490.567	0.00087	0.11368	0.00754	0.005227	0.006764	0.001614
27	0.974	49.746	2475.613	0.00293	0.09543	0.00912	0.005515	0.006219	0.002379
28	1.016	49.954	2496.434	0.00088	0.03273	0.00132	0.003593	0.003623	0.001444
29	1.014	50.044	2505.43	0.00173	0.03513	0.003105	0.003552	0.003751	0.001689
30	1.02	49.996	2500.64	0.0011	0.05213	0.0053	0.003595	0.004574	0.001471

4.4 that the assumption of the constant multivariate CVs is appropriate for the real-life dataset under consideration, as the plots of against $\bar{\mathbf{Y}}^T \bar{\mathbf{Y}}$ is constant for all the multivariate CV estimators. Specifically, there is no apparent correlation between $\hat{\gamma}^2$ and $\bar{\mathbf{Y}}^T \bar{\mathbf{Y}}$ hence, the notion of constant multivariate CV assumption is reasonable. Also, the results of the regression test provided in Table 4.5, which are conducted based on the null hypothesis which supports the assumption that the multivariate CV is constant, gives p -values which are higher than 0.05 significance level. Hence, the regression tests are not significant. Therefore, the null hypothesis is accepted, i.e., the assumption that the multivariate CV holds. This

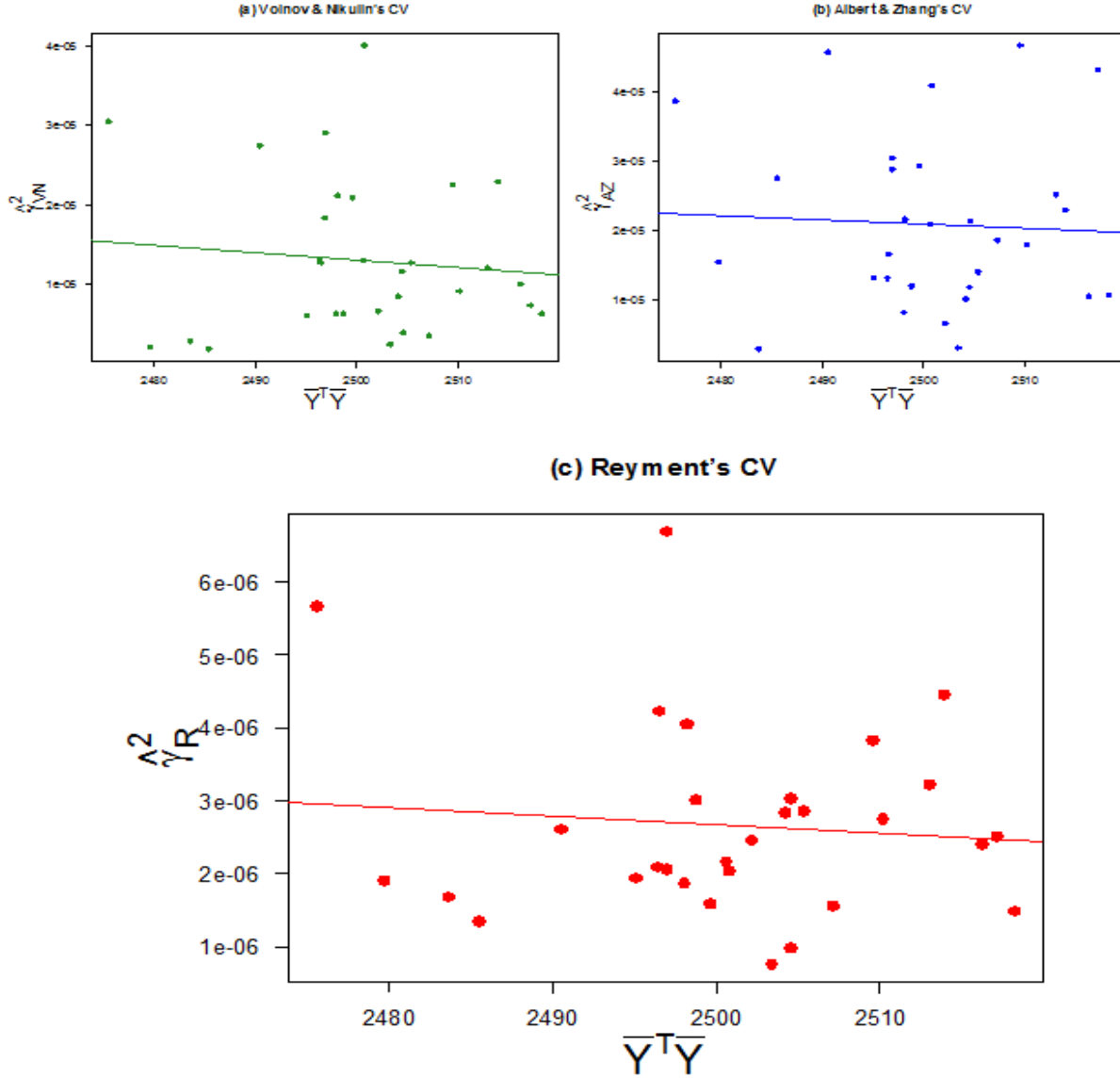


Figure 4.4: Scatter plot of the square multivariate CV statistics $\hat{\gamma}_{RR}$, $\hat{\gamma}_{VN}$, and $\hat{\gamma}_{AZ}$ against $\hat{\gamma}^2$ for the carbon fiber tubing data.

confirms that there is no evidence of the dependency of the multivariate CV on means.

The in-control multivariate CV is estimated by using all the three estimators i.e. $\hat{\gamma}_{VN} = 0.00335$, $\hat{\gamma}_{AZ} = 0.00437$ and $\hat{\gamma}_{RR} = 0.00159$. Based on these estimated CV's, the control limits were estimated and the control chart displays are presented in Figure 4.5. All the three charts indicate that the process CV is in-control as the monitoring statistic for all the charts are all within the control limits. To investigate the detection ability of the charts, shifts of size $\delta = 4$ was applied to the last m_1 samples of the Phase

Table 4.7: Regression test results for the carbon fiber tubing data

	Source of Variation	Dof	Sum of Squares	Mean Square	F-value	p-value
VN	Model	1	2.75E-11	2.75E-11	0.280	0.6010
	Error	28	2.75E-09	9.84E-11		
	Total	29	2.78E-09			
AZ	Model	1	1.12E-11	1.12E-11	0.069	0.7941
	Error	28	4.50E-09	1.61E-10		
	Total	29	4.50E-09			
R	Model	1	4.38E-13	4.38E-13	0.243	0.6261
	Error	28	5.05E-11	1.80E-12		
	Total	29	5.09E-11			

I dataset. We consider $m_1 = 6$ and 9, and the control chart displays are presented in Figure 4.6. As shown in these Figures, for both choices of m_1 , the multivariate MCV_R and MCV_{VN} charts show better detection ability as compared to MCV_{AZ} chart. Specifically, the multivariate MCV_{VN} chart gives the best performance in detecting shift in the process CV. This superiority of the MCV_{VN} chart is in accordance with the findings of Section 4.4, considering localized CV disturbances.

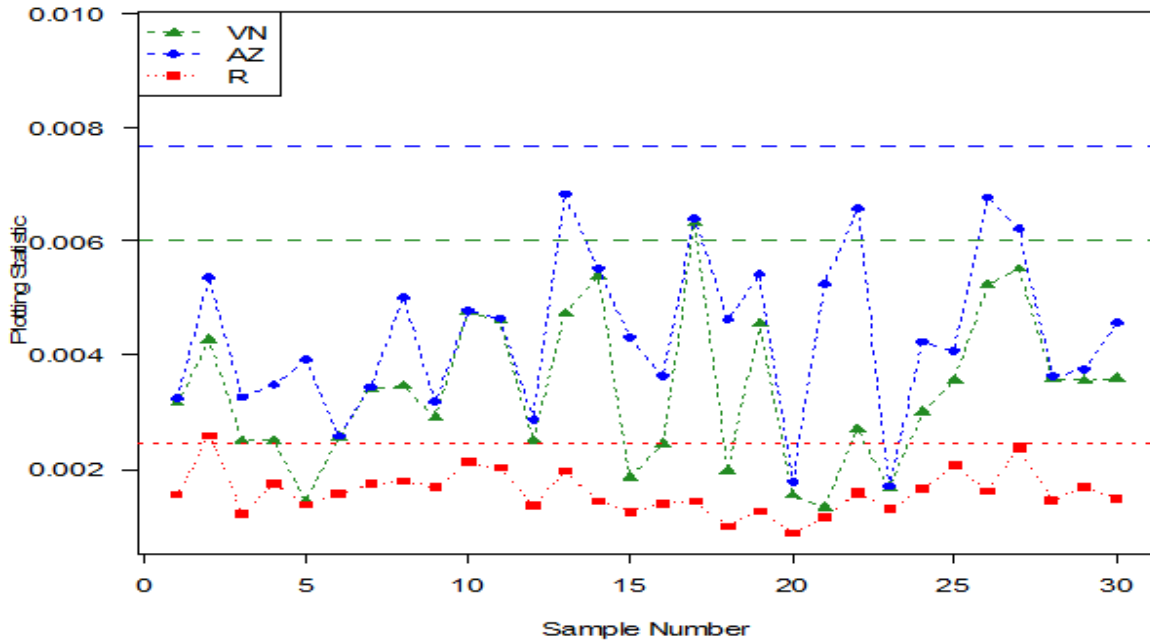


Figure 4.5: Control Chart plots for the three Multivariate CV charts for the carbon fiber tubing data, showing in-control state of the process.

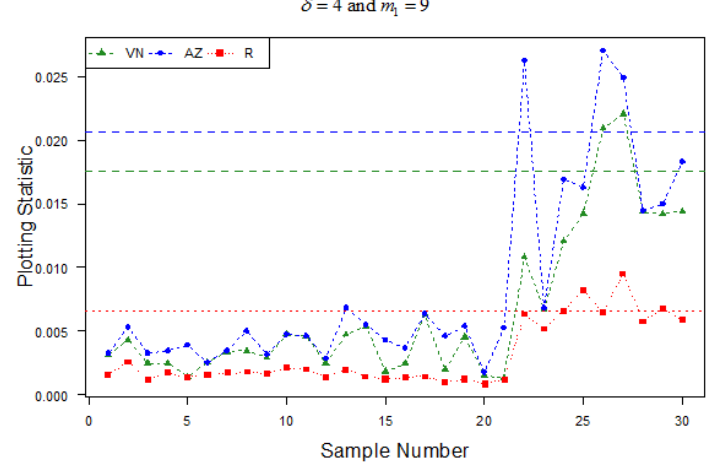
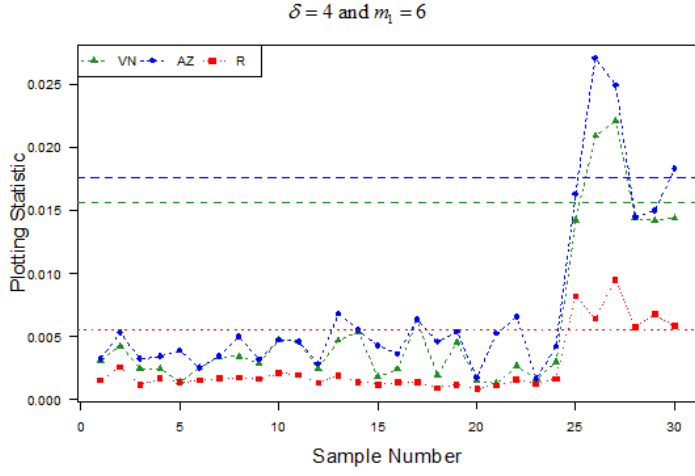


Figure 4.6: Control Chart plots for the three Multivariate CV charts in presence of localized CV disturbances for the carbon fiber tubing data.

4.6 Summary and conclusion

In this chapter, we investigated the choice of an appropriate CV control charting structure for efficient monitoring of multivariate CV in Phase I of SPC. We examined the performance of different multivariate CV charts under both localized CV and diffuse symmetric CV disturbance scenarios, using probability to signal (PTS) as a performance measure. It has been observed that, for all the multivariate CV charts, the PTS increases with an increase in sample size and the amount of shift in multivariate CV. The comparative analyses showed that the MCV_{AZ} chart exhibits the worst performance under both contaminated scenarios. The MCV_{VN} chart outperformed other competing charts in the presence of diffuse symmetric CV disturbances. The performance of the MCV_R chart is not too bad compared to the MCV_{VN} chart but significantly better than the MCV_{AZ} chart. On the other hand, the MCV_R chart appears to be the best choice in the presence of localized CV disturbances, as compared to both MCV_{VN} and MCV_{AZ} charts. The superiority of these charts is also demonstrated with a real-life example using inner diameter and length of carbon tubes. We recommend the use of MCV_R chart in Phase I of SPC due to the fact that the MCV_R chart is performing best for the detection of localized CV disturbances and its performance is very close to MCV_{VN} chart when it comes to detection of diffuse symmetric disturbances.

This study will help quality practitioners to choose an efficient multivariate CV control chart in Phase I.

Chapter 5

A new multivariate control chart based on homogeneously weighted moving average

”This is the peer reviewed version of the following article: “Adegoke, N. A., Abbasi, S. A., Smith, A. N., Anderson, M. J., & Pawley, M. D. (2019). A Multivariate Homogeneously Weighted Moving Average Control Chart. IEEE Access, 7, 9586-9597”, which has been published in final form at <https://doi.org/10.1109/ACCESS.2019.2891988>. This article may be used for non-commercial purposes in accordance with IEEE Terms and Conditions for Use of Self-Archived Versions.”

This chapter presents a multivariate homogeneously weighted moving average (MHWMA) control chart for monitoring a process mean vector. The MHWMA control chart statistic gives a specific weight to the current observation, and the remaining weight is evenly distributed among the previous observations. We present the design procedure and compare the average run length (ARL) performance of the proposed

chart with multivariate Chi-square, multivariate EWMA, and multivariate CUSUM control charts. The ARL comparison indicates superior performance of the MHWMA chart over its competitors, particularly for the detection of small shifts in the process mean vector. Examples are also provided to show the application of the proposed chart.

5.1 Introduction

Rapid developments in data-acquisition in industry have led to increased interest in the joint monitoring of several related process parameters (Bersimis et al., 2007). As a result, multivariate process control (MPC) methodology, in which several related process parameters are jointly monitored (Seif et al., 2011), is one of the most rapidly developing areas in statistical process control (SPC). Several MPC tools that use the relationships among the variables to provide efficient monitoring schemes for identifying any changes in the quality of the products have been proposed. These tools are capable of giving information as to when the process is in-control, provide diagnostic procedures for out-of-control situations, and are able to provide guidance on the overall process when it is out-of-control (Jackson, 1991). They are currently used in a range of scientific and technological application domains, including health-related monitoring, quality improvements, ecological monitoring, spatiotemporal surveillance, and profile monitoring (Montgomery, 2009).

MPC tools are generally applied in two monitoring phases (Montgomery, 2009). In Phase I, a historical reference sample is analysed to establish the values and stability of process parameters while in the in-control state. If the in-control parameter values are unknown, the data from Phase I are used to estimate these values and their control limits (Abbasi and Adegoke, 2018). In Phase II, the process parameters are monitored and checked for departure from the in-control state. If Phase II values (or statistics) remain inside the in-control Phase I limits, the process is believed to be in control; if they go outside the control

limits, this indicates that the process may be out-of-control and remedial actions are triggered.

Hotelling (1947) was the first to propose and employ a multivariate process control tool; his χ^2 statistic represented the weighted Mahalanobis distance of the sample point from the center of the cloud and is known as the multivariate χ^2 control chart. This chart signals whenever the χ^2 values obtained from the process variable are greater than the chart's control limit $h = \chi_{p,\alpha}^2$ (where $\chi_{p,\alpha}^2$ is the α^{th} upper percentage point of the chi-square distribution and p is the number of quality characteristics being monitored). The multivariate χ^2 chart is a memoryless-type chart that uses only the most current process information and disregards any previous observations, and very efficient in detecting large shifts in the process mean vector.

To increase the sensitivity of the multivariate process control tool for the detection of small-to-moderate shifts in the process mean vector, different multivariate memory-type tools that use information from both the current and previous process observations have been proposed. For example, Crosier (1988) and Pignatiello and Runger (1990) proposed different possible multivariate extensions of the univariate cumulative sum (CUSUM) chart proposed by Page (1961). The multivariate exponentially weighted moving average (EMWA) control chart proposed by Lowry et al. (1992) is a multivariate extension of the univariate EWMA chart proposed by Roberts (1959). The memory-type charts are particularly effective for individual-observation monitoring (Montgomery, 2009).

In this chapter, we propose a new memory-type multivariate charting procedure, namely, the multivariate homogeneously weighted moving average (MHWMA) control chart. Like other memory-type charts, MHWMA uses the current observation and past observations. However, previous methods allocate equal weight across the observations, including the current one. With our proposed MHWMA method, the weight of the current observation can be specified, with the remaining weight then allocated equally across previous observations. We will show that this can provide more efficient monitoring of small shifts in the process mean vector, when compared to other memory-type multivariate charting procedures.

The remainder of this article is organised as follows. A review of the design structures of the multivariate exponentially weighted moving average (MEWMA) chart by Lowry et al. (1992), the multivariate cumulative sum #1 (MCI) chart by Pignatiello and Runger (1990), and the multivariate cumulative sum (MCUSUM) chart by Crosier (1988), respectively, are provided in Section 5.2. The design of the MHWMA chart is discussed in Section 5.3, and the run length performance of the chart is evaluated in Section 5.4. The ARL comparisons of the MHWMA chart with that of the χ^2 chart, MEWMA chart, MCUSUM chart, and MCI chart, respectively, are provided in Section 5.5. Illustrative examples concerning the application of the proposed MHWMA chart are given in Section 5.6. Finally, conclusions and directions for future work are presented in Section 5.7.

In Appendix B.1, we derive the covariance matrix of the vector of HWMAs used with the MHWMA procedure. This matrix is used in Section 5.3 to obtain the MHWMA control-chart statistic. In Appendix B.2, we provide the proof of the dependency of the ARL performance of the MHWMA chart on the mean vector and covariance matrix only through the non-centrality parameter.

5.2 Literature review: the memory-type control charts

Suppose we have $p \times n$ independently and identically distributed multivariate normal random variables $\mathbf{Y}_1, \mathbf{Y}_2, \dots$, with mean vector $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$. For monitoring the mean vector ($\boldsymbol{\mu}_0$) of an individual-observation (i.e., $n = 1$), the design structures of the memory-type charts are briefly described below:

5.2.1 The MCUSUM chart

Crosier (1988) proposed two multivariate CUSUM charts. The one with better ARL performance obtains the CUSUM vector directly from the multivariate observation, and the MCUSUM vectors for the observed

vector y_i are given as:

$$C_i = [(S_{i-1} + y_i - \mu_0)' \Sigma_0^{-1} (S_{i-1} + y_i - \mu_0)]^{1/2} \quad (5.1)$$

where

$$S_i = \mathbf{0} \quad \text{if } C_i \leq k$$

$$S_i = (S_{i-1} + y_i - \mu_0)(1 - k/C_i) \quad \text{if } C_i > k$$

$S_0 = \mathbf{0}$ and $k > 0$. The MCUSUM control chart signals when $T_i^2 = [S_i' \Sigma_0^{-1} S_i] > h$.

5.2.2 The MCI chart

Two directionally invariant multivariate CUSUM charts were proposed by Pignatiello and Runger (1990); the one with better ARL performance is the MCI chart. Here, the CUSUM vectors for the observed vector y_i are given as:

$$C_i = \sum_{j=i-n_i+1}^i (y_j - \mu_0) \quad (5.2)$$

$$T_i = \max \left\{ \sqrt{C_i' \Sigma_0^{-1} C_i} - kn_i, 0 \right\}$$

and

$$n_i = \begin{cases} n_{i-1} + 1, & \text{if } T_{i-1} > 0 \\ 1, & \text{if otherwise} \end{cases}$$

where n_i ($i = 1, 2, \dots$), is interpreted as the number of subgroups up to the most recent cumulative sum statistic. The MCI control chart signals when $T_i > h$, for positive values of $h > 0$ and $k > 0$. The parameters of the MCUSUM and MCI charts, k and h , are chosen to give the desired in-control ARL

performance of the chart Crosier (1988); Pignatiello and Runger (1990).

5.2.3 The MEWMA chart

The MEMWA control chart, proposed by Lowry et al. (1992), is a multivariate extension of the EWMA chart. It is a memory-type method that accumulates information from previous observations. The MEWMA statistics for the observed vector \mathbf{y}_i are given as:

$$\mathbf{P}_i = r\mathbf{y}_i + (1 - r)\mathbf{P}_{i-1} \quad (5.3)$$

The use of small values for the smoothing parameter increases the power of the control chart and, if $r = 1$, the chart is identical to the memoryless control chart based on Hotelling's T^2 .

The MEWMA chart gives an out-of-control signal when:

$$T_i^2 = (\mathbf{P}_i - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_{\mathbf{P}_i}^{-1} (\mathbf{P}_i - \boldsymbol{\mu}_0) > h \quad (5.4)$$

where h and r are chosen to achieve a desired in-control performance measure (such as a desired value of in-control ARL), and $\boldsymbol{\Sigma}_{\mathbf{P}_i}$ is the covariance matrix at time point i . Lowry et al. (1992) provided two alternative forms of $\boldsymbol{\Sigma}_{\mathbf{P}_i}$: the exact covariance matrix is given as:

$$\boldsymbol{\Sigma}_{\mathbf{P}_i} = \frac{r[1 - (1 - r)^{2i}]}{2 - r} \boldsymbol{\Sigma}_0 \quad (5.5)$$

and the asymptotic covariance matrix is given as:

$$\boldsymbol{\Sigma}_{\mathbf{P}} = \frac{r}{2 - r} \boldsymbol{\Sigma}_0 \quad (5.6)$$

The MEWMA, MCUSUM and MCI charts are directionally invariant charts; the ARL performance of the charts depend on $\boldsymbol{\mu}_0$ and $\boldsymbol{\Sigma}_0$, only through the non-centrality parameter given as:

$$\delta = \sqrt{(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)} \quad (5.7)$$

where $\boldsymbol{\mu}_1$ is the mean vector for the out-of-control process.

Several enhancements of these memory-type control charts in detecting small-to-moderate shifts have been proposed in SPC and related literature. For example, Hawkins, Douglas M and Maboudou-Tchao (2007) proposed a self-starting MEWMA control charting for monitoring the process mean vector. Also, a self-starting control chart for multivariate individual observations monitoring was proposed by Sullivan and Jones (2002). Kramer and Schmid (1997) proposed EWMA charts for multivariate time series observation monitoring. Ngai and Zhang (2001) proposed a MCUSUM control chart based on projection pursuit. Park and Jun (2015) investigated a MEWMA control chart via multiple testing. Qiu and Hawkins (2003) proposed a nonparametric MCUSUM procedure for detecting shifts in all directions. Qiu and Hawkins (2001) proposed a rank-based MCUSUM Procedure. A multivariate sign EWMA control chart was proposed by Tsung and To (2012). A cumulative sum control charts for monitoring the covariance matrix (Chan, Lai K and Zhang, 2001). A MEWMA control chart that can handle a non-constant smoothing parameter of the chart was proposed by Yumin (1996). An adaptive multivariate CUSUM control chart for signaling a range of location shifts was proposed by Wang and Huang (2016). The performance of multivariate memory-type control charts with estimated parameters are investigated by Jones et al. (2001); Mahmouda and Maravelakisb (2010); Aly et al. (2016); Champ and Jones-Farmer (2007); Mahmoud and Maravelakis (2013).

5.3 The multivariate homogeneously weighted moving average (MHWMA) control chart

To increase the sensitivity of the memory-type charts given in Section 5.2 in monitoring small shifts in the process mean vector, we propose a MHWMA control chart. The MHWMA control chart statistic gives a specific weight to the current observation, and the remaining weight is evenly distributed among the previous observations. The monitoring statistic of the proposed MHWMA chart is defined as:

$$\mathbf{H}_i = \mathbf{W}\mathbf{y}_i + (\mathbf{I} - \mathbf{W})\bar{\mathbf{y}}_{i-1} \quad (5.8)$$

where, $i = 1, 2, \dots$, $\bar{\mathbf{y}}_{i-1}$ represents the sample average of the previous information up to and including the $i - 1$ observation, and $\bar{\mathbf{y}}_0 = \boldsymbol{\mu}_0$. \mathbf{W} is a $p \times p$ diagonal square matrix with smoothing or sensitivity parameters w_k , $k = 1, 2, \dots, p$, along the diagonal such that $0 < w_k \leq 1$. The matrix \mathbf{I} is a diagonal matrix of 1's. If the values of the smoothing parameter, which determine the weight of each prior observation, are equal across variables, then the MHWMA vector becomes:

$$\mathbf{H}_i = w\mathbf{y}_i + (1 - w)\bar{\mathbf{y}}_{i-1} \quad (5.9)$$

The MHWMA chart gives an out-of-control signal when

$$T_i^2 = (\mathbf{H}_i - \boldsymbol{\mu}_0)' \boldsymbol{\Sigma}_{\mathbf{H}_i}^{-1} (\mathbf{H}_i - \boldsymbol{\mu}_0) > h \quad (5.10)$$

Here, h and w are chosen to achieve a desired in-control ARL performance measure, and $\boldsymbol{\Sigma}_{\mathbf{H}_i}$ is the

covariance matrix at time point i . From Appendix A, we have

$$\Sigma_{H_i} = \begin{cases} w^2 \Sigma_0 & \text{if } i = 1 \\ w^2 \Sigma_0 + (1-w)^2 \frac{\Sigma_0}{i-1} & \text{if } i > 1 \end{cases} \quad (5.11)$$

The MHWMA chart is a directionally invariant chart. In Appendix B, we give a proof that shows the relationship between the ARL performance and the non-centrality parameter given in equation 5.7.

Special cases

- If $w = 1$, the monitoring statistic in equation (5.9) becomes:

$$H_i = y_i \quad (5.12)$$

and, Σ_{H_i} in equation (5.11) becomes:

$$\Sigma_{H_i} = \Sigma_0 \quad (5.13)$$

In this case, the MHWMA chart is identical to the memoryless χ^2 control chart, and we recommend monitoring either the χ^2 chart or the MHWMA chart (with $w = 1$).

- If $p = 1$, the monitoring statistic in equation (5.9) becomes:

$$H_i = wy_i + (1-w)\bar{y}_{i-1} \quad (5.14)$$

and, the variance of the monitoring statistic H_i in equation (5.14) becomes:

$$\sigma_{H_i} = \begin{cases} w^2 \sigma_0^2 & \text{if } i = 1 \\ w^2 \sigma_0^2 + (1-w)^2 \frac{\sigma_0^2}{i-1} & \text{if } i > 1 \end{cases} \quad (5.15)$$

where σ_0^2 is the variance of a normally distributed univariate random variable. In this case, we recommend monitoring the proposed chart by Abbas (2018).

- When $n > 1$, the vector \mathbf{y} in the plotting statistic for the MHWMA vector in equation (5.9) can be replaced by the average of the i th sample (i.e, $\bar{\mathbf{y}}$). Hence, the covariance structure of the MHWMA chart becomes

$$\Sigma_{H_i} = \begin{cases} \frac{w^2}{n} \Sigma_0 & \text{if } i = 1 \\ \frac{w^2}{n} \Sigma_0 + (1-w)^2 \frac{\Sigma_0}{n(i-1)} & \text{if } i > 1 \end{cases} \quad (5.16)$$

5.4 Performance evaluation

In this section, we evaluate the performance of the proposed MHWMA chart by using different run length characteristics such as the average run length and standard deviation of the run length (SDRL) distribution. ARL is the most commonly used performance measures for control chart procedures. The in-control ARL (denoted by ARL_0), is the average number of plotted samples until an out-of-control signal is detected by a control chart when the process is in control. The out-of-control ARL (denoted by ARL_1), is the average number of plotted points until an out-of-control signal is detected by a control chart when the process is out of control (Aldosari et al., 2018; Aslam et al., 2018; Aslam, 2018; Adegoke et al., 2017; Abbasi et al., 2013). It is generally desirable to have large values of ARL_0 and small values of ARL_1 for any control-chart setting. The SDRL measures the spread of the run length distribution (Human et al., 2011). Similarly, $SDRL_0$ and $SDRL_1$ can be defined.

Table 5.1: ARL Values for MHWMA Charts ($p = 2$).

	w												
	0.03		0.05		0.2		0.4		0.6		0.8		
δ	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	
0	199.26	202.01	199.35	182.57	202.99	186.06	201.63	192.36	200.54	201.62	200.39	199.06	
0.05	166.24	197.29	173.00	157.48	186.94	169.67	193.77	189.26	194.31	190.36	199.62	198.50	
0.10	115.64	132.81	129.55	119.99	154.16	136.29	177.88	173.09	184.99	185.98	192.07	194.5	
0.25	43.05	44.91	53.32	46.28	74.00	60.41	107.64	102.17	139.03	137.16	161.6	160.77	
0.50	15.80	14.3	19.87	15.56	27.50	19.5	41.84	36.4	64.86	62.32	91.91	90.24	
0.75	8.56	6.83	10.65	7.7	14.50	9.24	19.25	15.5	31.66	29.95	50.15	49.11	
1.00	5.70	4.09	6.91	4.52	9.23	5.31	10.94	7.96	16.54	14.56	27.24	26.23	
1.50	3.32	2.04	3.90	2.22	4.85	2.48	5.05	3.05	6.3	4.8	9.69	8.71	
2.00	2.32	1.4	2.67	1.49	3.19	1.49	3.13	1.62	3.41	2.2	4.54	3.63	
2.50	1.74	1.06	1.99	1.16	2.35	1.12	2.22	1.05	2.26	1.2	2.66	1.79	
3.00	1.37	0.79	1.53	0.9	1.81	0.9	1.7	0.76	1.7	0.79	1.86	1.07	
5.00	1.00	0.08	1.01	0.09	1.03	0.17	1.03	0.16	1.03	0.18	1.03	0.17	
h													
5.40			6.79			10.19			10.56			10.61	
												10.62	

The results are based on 10^5 Monte Carlo simulations, and δ denotes the shift size (given in equation (5.7)). The appropriate values of h are also obtained using simulation. Table 5.1 reports the ARL and SDRL results for the case when $p = 2$ at varying levels of smoothing parameter (w) and shift (δ). See Appendix C.2 for the *R code* used in obtaining the ARL and SDRL of the proposed chart. The chart's parameters in Table 5.1 are chosen to fix the ARL_0 at 200. Visual representation of the logarithm of the ARL values in Table 5.1 are also provided in Figure 5.1. From the reported results in Table 5.1 (and/or Figure 5.1), we observe that:

- Smaller values of w are more effective in detecting shifts in the mean vector. Specifically, the use of small values for the smoothing parameter increases the power of the MHWMA control chart.
- The proposed MHWMA chart is ARL unbiased, i.e., for any combinations of h and w , the ARL_1 values from the chart are always lesser than the ARL_0 .
- The higher the ARL values of the chart, the higher the SDRL value as well.
- It is apparent that both ARL and SDRL decrease as the size of the shift increases. This indicates that larger shifts can be detected quickly and will result in a smaller spread in the run length

distribution.

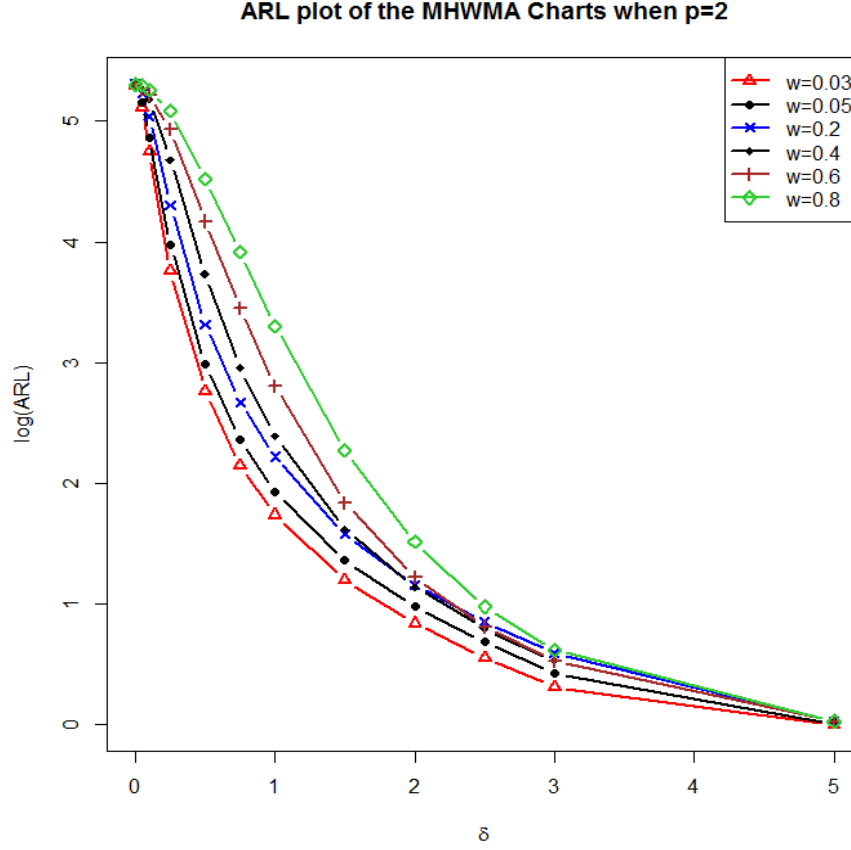


Figure 5.1: Plot of the logarithms of the ARL values given in Table 5.1.

Tables 5.2-5.3 report the ARL and SDRL results for the case when $w = 0.1$ but with varying levels of p (i.e., $p = 2, 3$, and 4), and δ . The values shown for parameter h , in each case, are chosen such that the ARL_0 is fixed at 50, 100, 500, or 1,000, respectively. We used $w = 0.10$, because small values of w are effective at detecting small shifts in the mean vector. From the reported results in Tables 5.2-5.3, we observe that:

- The ARL and SDRL performance of the chart depend on the number of quality characteristics (p).

Specifically, the performance of the chart increases with the small value of p .

- The logarithm of the in-control ARL is very close to a linear function of the chart's upper limits.

This property of the MHWMA chart can be used to approximate the appropriate value of the chart's control limits for other in-control ARL 's.

- Larger shifts are detected quickly and result in a smaller spread in the run-length distribution.

Table 5.2: ARL values for MHWMA charts ($w = 0.1$).

δ	$p = 2$				$p = 3$				$p = 4$			
	5.27	7.01	11.52	13.31	h				8.60	10.8	16.01	17.98
0	51.04	102.17	500.23	1000.53	51.01	103.07	502.15	997.44	50.07	102.32	498.68	1001.21
0.05	48.99	95.51	419.86	753.6	49.46	96.37	422.95	789.64	49.65	97.48	434.69	790.14
0.1	44.75	81.68	282.59	439.03	45.06	84.3	297.86	475.59	45.22	85.66	313.6	497.17
0.25	27.74	43.66	96.55	123.65	29.54	47.15	105.34	137.94	30.03	48.93	113.53	150.3
0.5	13.31	18.56	33.7	40.35	14.38	20.66	37.18	44.56	15.44	22.34	40.61	47.83
0.75	7.81	10.39	17.39	20.52	8.51	11.56	19.17	22.52	9.2	12.44	20.87	24.13
1	5.27	6.84	10.86	12.49	5.83	7.59	11.98	13.79	6.2	8.12	12.9	14.74
1.5	3.08	3.85	5.74	6.43	3.42	4.21	6.23	6.97	3.6	4.51	6.65	7.51
2	2.15	2.63	3.78	4.2	2.38	2.87	4.08	4.53	2.49	3.09	4.32	4.8
2.5	1.64	1.96	2.8	3.1	1.77	2.15	3.02	3.35	1.9	2.27	3.21	3.51
3	1.3	1.51	2.17	2.41	1.41	1.66	2.34	2.61	1.49	1.78	2.5	2.77
5	1	1.01	1.06	1.1	1.01	1.01	1.09	1.15	1.01	1.03	1.12	1.2

Table 5.3: SDRL values for MHWMA charts ($w = 0.1$).

δ	$p = 2$				$p = 3$				$p = 4$			
	5.27	7.01	11.52	13.31	h				8.60	10.8	16.01	17.98
0	47.77	85.58	415.84	869.58	48.22	87.45	423.09	872.89	48.39	88.1	428.78	895.68
0.05	45.55	80.29	344.37	648.49	46.31	82.71	349.15	688.85	46.86	83.06	364.29	678.68
0.1	41.98	68.54	224.9	356.51	43.11	71.44	237.45	386.67	43.22	72.95	251.34	408.2
0.25	25.48	35.21	67.11	82.15	26.96	38.13	72.29	92.24	27.35	39.51	77.69	100.23
0.5	11.15	13.9	20.31	23.13	11.88	15	22.06	25.13	12.7	16.1	23.85	26.82
0.75	5.98	7.13	9.81	10.92	6.47	7.79	10.59	11.72	6.94	8.37	11.3	12.28
1	3.71	4.36	5.81	6.29	4.04	4.79	6.28	6.8	4.35	5.13	6.66	7.17
1.5	1.89	2.16	2.72	2.9	2.08	2.32	2.9	3.14	2.16	2.45	3.09	3.32
2	1.28	1.42	1.64	1.73	1.37	1.5	1.73	1.81	1.43	1.57	1.8	1.9
2.5	0.96	1.1	1.25	1.26	1.03	1.17	1.27	1.3	1.1	1.21	1.3	1.29
3	0.68	0.85	1.07	1.09	0.77	0.93	1.09	1.11	0.84	0.98	1.11	1.11
5	0.05	0.09	0.28	0.37	0.09	0.14	0.35	0.45	0.11	0.19	0.42	0.52

5.5 Average run length comparisons

In this section, the (zero-state) ARL performance of the MHWMA chart is compared with that of the χ^2 chart, the MCUSUM chart by Crosier (1988), the MCI chart by Pignatiello and Runger (1990), and the MEWMA chart by Lowry et al. (1992). Since, the MEWMA, the MCUSUM, the MCI and the Hotelling's χ^2 charts are all directional invariant; these charts can be compared with each other and with the proposed MHWMA chart. We consider both the time-varying and the asymptotic limits MEWMA control chart.

The ARL values of the charts are presented in Tables 5.4 to 5.9, for $p = 2, 3, 4, 5, 10$ and 20 , respectively. To allow reasonable comparisons of the proposed chart with the other charts, each chart is designed to give ARL_0 of approximately 200. We observed from Tables 5.4 to 5.9 that:

- The Hotelling's χ^2 chart, the MCUSUM chart, the MCI chart, and the MEWMA chart based on the asymptotic covariance structure (given in equation (5.6)), respectively, are all inferior to the proposed MHWMA chart (i.e., the MHWMA chart resulted in smallest values of the ARL_1) across all shifts.
- The simulation results show that the MHWMA chart detects shifts more rapidly than the MEWMA chart based on the exact covariance structure when $\delta \leq 0.5$. However, the ARL performance the MEWMA chart (given in equation (5.5)) is superior to the ARL performance of the proposed chart when there is a moderate-to-large shift in the mean vector. Specifically, the ARL_1 value of MEWMA chart based on the varying limit is smaller than the proposed chart when $\delta > 0.5$ is considered.

5.6 Illustrative example

In this section, we provide a couple of examples for illustrating the application of the proposed MHWMA chart. The first example is based on a simulated dataset following Crosier (1988), whereas, the second

Table 5.4: ARL comparisons for $p = 2$.

	χ^2 ;	MCI	MCUSUM;	MEWMA (5.5);	MEWMA (5.6);	MHWMA;
δ	$h = 10.60$	$k_1 = 0.50,$ $h = 4.75$	$k_2 = 0.5,$ $h = 5.50$	$r = 0.1, 0$ $h = 8.79$	$r = 0.1,$ $h = 8.66$	$w = 0.1,$ $h = 8.965$
0.00	201.08	202.27	201.34	202.01	200.54	202.64
0.05	198.13	190.92	192.48	187.92	190.68	181.90
0.10	194.47	169.74	166.02	159.35	163.79	144.53
0.25	171.97	91.65	83.85	73.69	77.20	64.12
0.50	117.39	31.40	29.91	25.08	28.02	24.94
0.75	71.04	15.00	15.11	12.62	15.21	13.49
1.00	41.95	9.44	9.92	7.76	10.14	8.61
1.50	15.78	5.26	5.78	4.06	6.09	4.63
2.00	6.80	3.69	4.11	2.60	4.41	3.15
2.50	3.56	2.90	3.24	1.90	3.50	2.32
3.00	2.14	2.42	2.69	1.50	2.94	1.78
5.00	1.03	1.58	1.82	1.01	1.97	1.02

Table 5.5: ARL comparisons for $p = 3$.

	χ^2	MCI	MCUSUM	MEWMA (5.5);	MEWMA (5.6);	MHWMA
δ	$h = 12.85$	$k_1 = 0.50$ $h = 5.48$	$k_2 = 0.5$ $h = 6.88$	$r = 0.10$ $h = 10.97$	$r = 0.1$ $h = 10.79$	$w = 0.1$ $h = 11.09$
0.00	200.90	198.29	199.07	199.38	200.08	198.91
0.05	198.49	192.56	189.36	189.23	190.41	182.28
0.10	196.50	173.47	165.69	164.35	164.96	148.01
0.25	179.66	99.52	86.58	83.57	85.22	70.22
0.50	130.17	34.17	31.70	29.04	32.07	27.26
0.75	83.78	16.33	16.78	14.23	17.00	14.84
1.00	52.27	10.08	11.18	8.78	11.20	9.47
1.50	19.94	5.70	6.74	4.48	6.72	5.06
2.00	8.81	4.04	4.84	2.88	4.85	3.41
2.50	4.42	3.18	3.81	2.07	3.82	2.51
3.00	2.54	2.65	3.18	1.61	3.20	1.96
5.00	1.05	1.80	2.03	1.03	2.04	1.03

example is based on the bimetal dataset given in Santos-Fernandez (2012).

5.6.1 Simulated example

The dataset (see Table 5.10) is from a similar example given by Crosier (1988), and also used for illustration in Lowry et al. (1992). The data consists of 10 observations, the mean is in-control at $\mu_0 = (0, 0)$ for the first five observations and out-of-control at $\mu_0 = (1, 2)$ for the last five observations. This example is

Table 5.6: ARL comparisons for $p = 4$.

		MCUSUM;	MEWMA; (5.5);	MEWMA (5.6);	MHWMA;
	χ^2 ;	$k_2 = 0.5$,	$r = 0.10$,	$r = 0.1$,	$w = 0.1$,
δ	$h = 14.86$	$h = 8.15$	$h = 12.93$	$h = 12.73$	$h = 13.11$
0.00	199.67	198.82	200.35	201.31	202.48
0.05	199.22	189.39	192.01	192.38	185.14
0.10	196.49	166.35	170.33	170.00	157.64
0.25	182.25	86.83	90.02	94.14	74.38
0.50	139.90	33.24	31.94	35.25	30.01
0.75	94.32	18.28	15.54	18.40	16.32
1.00	61.12	12.46	9.45	12.05	10.21
1.50	24.27	7.57	4.83	7.23	5.42
2.00	10.77	5.47	3.07	5.17	3.62
2.50	5.16	4.33	2.21	4.09	2.71
3.00	2.93	3.61	1.71	3.41	2.09
5.00	1.07	2.20	1.04	2.11	1.05

Table 5.7: ARL comparisons for $p = 5$.

		MCI;	MCUSUM;	MEWMA (5.5);	MEWMA (5.6);	MHWMA;
	χ^2 ;	$k_1 = 0.50$,	$k_2 = 0.5$,	$r = 0.10$,	$r = 0.1$,	$w = 0., 1$
δ	$h = 16.75$	$h = 6.81$	$h = 9.46$	$h = 14.74$	$h = 14.56$	$h = 14.92$
0.00	201.27	204.29	200.10	201.17	201.09	201.25
0.05	199.91	194.57	191.99	192.94	193.66	187.90
0.10	198.00	178.31	175.26	172.87	174.92	159.04
0.25	184.88	109.22	91.95	95.63	99.13	77.97
0.50	143.82	38.74	35.81	34.40	38.35	31.38
0.75	102.37	17.72	19.95	16.71	19.83	17.08
1.00	68.25	11.04	13.71	10.12	12.91	10.78
1.50	28.44	6.39	8.47	5.16	7.65	5.79
2.00	12.27	4.60	6.16	3.24	5.47	3.85
2.50	6.00	3.64	4.87	2.34	4.31	2.82
3.00	3.35	3.03	4.05	1.80	3.60	2.23
5.00	1.10	2.01	2.53	1.06	2.20	1.08

illustrative of a moderate-to-large shift in the process mean vector, as the size of δ (in equation (5.7)) is approximately 2.65.

The first two columns of Table 5.10 give the sample of bivariate observations for the random variables Y_1 and Y_2 . The columns H_1 and H_2 are the corresponding values of the MHWMA vector as provided in equation (5.9) using $w = 0.10$. The T^2 values obtained from equation (5.10) are given in the last column. For a fair comparison, the control limits were selected to give the desired ARL_0 of 200 for all the charts

Table 5.8: ARL comparisons for $p = 10$.

		MCUSUM;	MEWMA (5.5);	MEWMA (5.6);	MHWMA;
δ	χ^2 ; $h = 25.19$	$k_2 = 0.5$, $h = 14.90$	$r = 0.10$, $h = 22.91$	$r = 0.1$, $h = 22.67$	$w = 0.1$, $h = 23.08$
0.00	200.03	198.63	195.71	204.11	199.09
0.05	199.15	192.01	194.58	197.11	190.02
0.10	198.4	170.55	173.78	182.80	162.23
0.25	190.45	96.34	114.45	117.35	89.52
0.50	162.01	42.89	45.20	47.94	37.53
0.75	124.50	25.96	21.47	24.94	20.86
1.00	92.80	18.62	12.60	15.85	13.05
1.50	44.70	11.92	6.38	9.19	6.87
2.00	20.60	8.80	3.97	6.57	4.52
2.50	9.90	7.02	2.78	5.15	3.36
3.00	5.20	5.86	2.13	4.28	2.64
5.00	1.24	3.63	1.12	2.69	1.19

Table 5.9: ARL comparisons for $p = 20$.

		MCUSUM;	MEWMA (5.5);	MEWMA (5.6);	MHWMA;
δ	χ^2 ; $h = 40.00$	$k_2 = 0.5$, $h = 24.70$	$r = 0.10$, $h = 37.32$	$r = 0.1$, $h = 37.01$	$w = 0.1$, $h = 37.59$
0.00	202.01	199.21	198.61	200.78	202.12
0.05	199.98	193.02	194.59	197.51	193.61
0.10	198.73	175.36	185.65	187.59	173.59
0.25	193.11	109.37	130.76	135.51	102.77
0.50	173.05	55.99	58.31	63.09	46.24
0.75	145.62	36.65	27.89	32.27	25.60
1.00	116.96	27.18	16.32	20.16	16.34
1.50	66.53	17.94	7.99	11.28	8.50
2.00	34.46	13.50	4.90	8.00	5.47
2.50	17.26	10.81	3.40	6.26	4.01
3.00	9.07	9.05	2.59	5.20	3.13
5.00	1.58	5.57	1.29	3.19	1.44

using $w = 0.10$. A plot of the MCUSUM chart with the same ARL_0 of 200, given by Crosier (1988) (also reproduced in Figure 5.2), signals after the tenth observation. Plots of the MEWMA charts based on the exact and asymptotic limits of the same in-control ARL, given by Lowry et al. (1992), signals after the ninth and tenth observation, respectively. The plot of the MCI and MHWMA charts also signal an out-of-control situation after the tenth observation.

Table 5.10: The simulated dataset.

n	Observation		MHWMA vector		T^2
	Y_1	Y_2	H_1	H_2	
1	-1.19	0.59	-0.12	0.06	3.29
2	0.12	0.9	-1.06	0.62	3.52
3	-1.69	0.4	-0.65	0.71	4.47
4	0.3	0.46	-0.80	0.61	7.15
5	0.89	-0.75	-0.46	0.45	3.97
6	0.82	0.98	-0.20	0.39	2.07
7	-0.3	2.28	-0.14	0.62	4.47
8	0.63	1.75	-0.07	0.80	7.45
9	1.56	1.58	0.11	0.90	8.71
10	1.46	3.05	0.26	1.12	13.85*
h					8.965

* Out-of-control signal

5.6.2 Bimetal thermostat dataset example

For the second example, we used the bimetal thermostat dataset taken from Santos-Fernandez (2012). The dataset contains measurements of the deflection, curvature, resistivity, and hardness for each of the low and high-expansion sides of brass and steel bimetal thermostats (Adegoke et al., 2018b). The process was employed in Phase I and Phase II, and data from the process at each phase consisted of sample size $m = 28$, and with $p = 5$ variables. The Phase I process is used to study a historical reference sample, which involves establishing the in-control state and evaluating the process stability to ensure that the reference sample is representative of the process. After this, the process parameters μ_0 and Σ_0 , are estimated from Phase I, and control chart limits are obtained to be used in Phase II. The Phase II aspect involves on-line monitoring of the process. In essence, any shift in the process needs to be detected quickly in Phase II, so that corrective actions can be taken at an early stage.

The estimated mean vector ($\hat{\mu}_0$) and covariance matrix ($\hat{\Sigma}_0$) are shown below.

$$\hat{\mu}_0 = (21.01607, 40.01607, 15.19214, 22.02393, 26.01214)$$

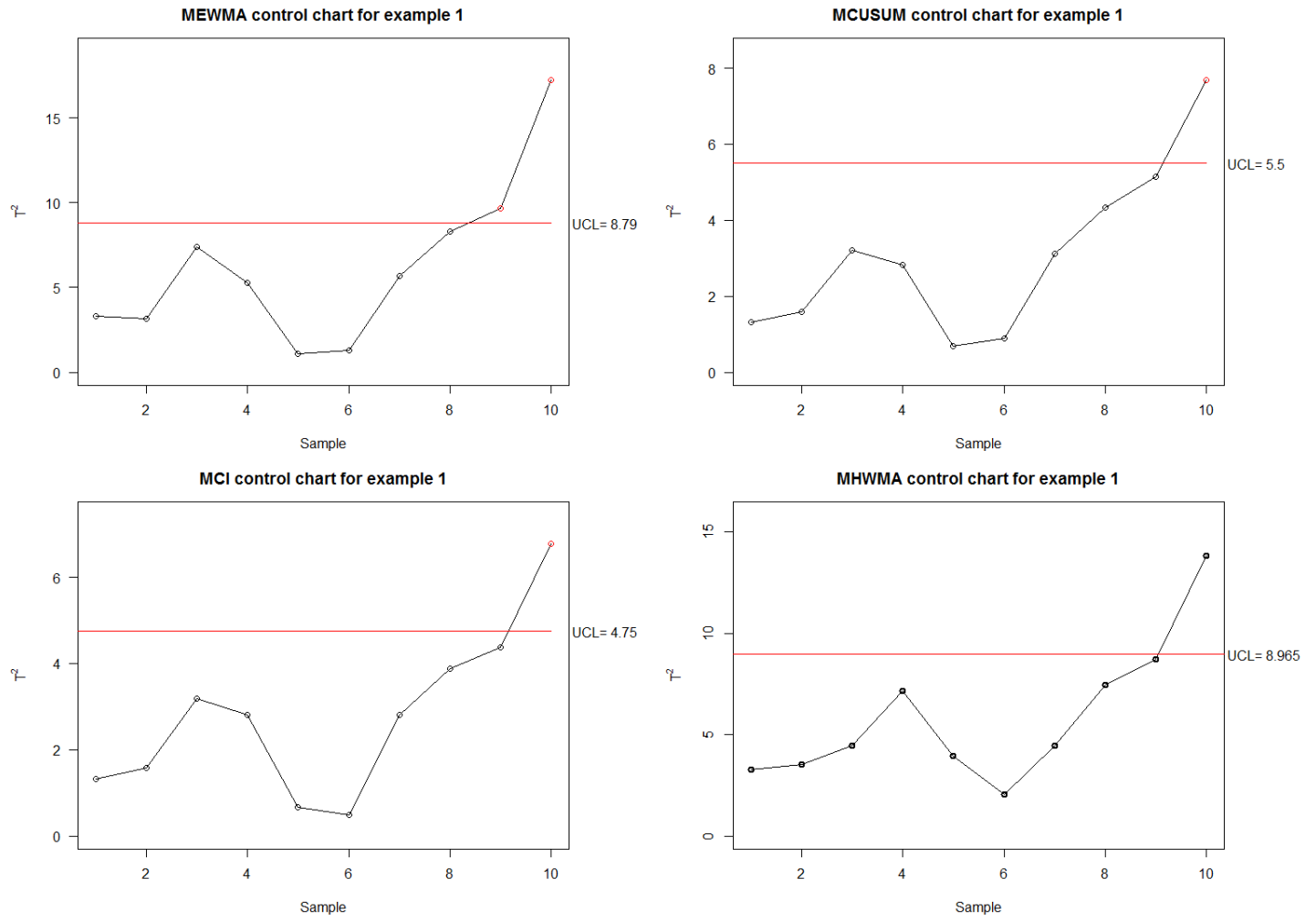


Figure 5.2: Plots of the memory-type charts of the simulated dataset

$$\hat{\Sigma}_0 = \begin{pmatrix} D & C & R & HL & HH \\ \begin{pmatrix} 0.091877 & 0.025443 & 0.037909 & 0.027931 & 0.026753 \\ 0.025443 & 0.018543 & 0.026342 & 0.016131 & 0.016998 \\ 0.037909 & 0.026342 & 0.106284 & 0.016439 & 0.023377 \\ 0.027931 & 0.016131 & 0.016439 & 0.05444 & 0.011088 \\ 0.026753 & 0.016998 & 0.023377 & 0.011088 & 0.021477 \end{pmatrix} & D \\ & C \\ & R \\ & HL \\ & HH \end{pmatrix}$$

Considering these estimates as the known parameters, we generated 20 Phase II observations from

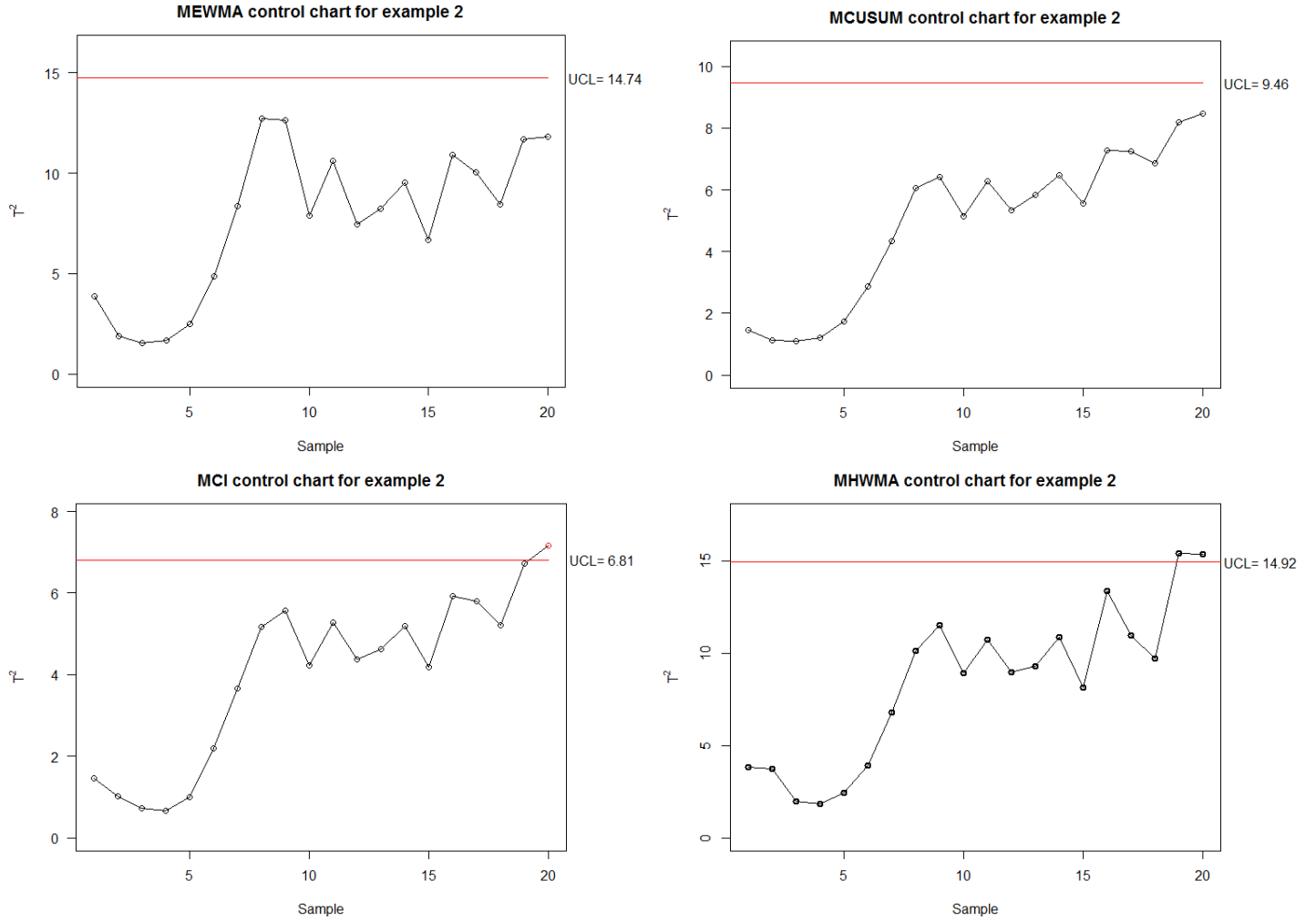
a multivariate normal distribution with mean $\boldsymbol{\mu}_1$ and covariance matrix $\hat{\boldsymbol{\Sigma}}_0$, such that the size of δ (in equation (5.7)) is approximately 0.087, which is a small shift in the mean vector. Specifically, we used, $\boldsymbol{\mu}_1 = (21.12, 40.12, 15.29, 22.12, 26.11)$. The inspiration for generating data in such manner is taken from Sanusi et al. (2017b); Singh, Ravindra and Mangat (2013). The simulated bimetal Phase II data is given in Table 5.11.

The first five columns of Table 5.11 give the sample number (n) and the observations of the random variables: the deflection (D), curvature (C), resistivity (R), Hardness low side (HL), and Hardness high side (HS). The columns H_1, H_2, H_3, H_4 , and H_5 are the corresponding values of the MHWMA vector from equation (5.9) with $w = 0.10$. The T^2 values obtained from equation (5.10) are given in the last column. The values of the control limits and w were used to give an ARL_0 of 200. The control limits are obtained from Table 5.7 for all of the charts. The MEWMA chart with time-varying structure and the MCUSUM chart failed to detect the out-of-control signal (see Figure 5.3). The MCI chart detected the signal after the twentieth observation, while the MHWMA chart detected the shift in the mean vector after the nineteenth observation.

Table 5.11: Simulated bimetal Phase II dataset.

n	Observation					MHWMA vector					T^2
	D	C	R	HL	HH	H_1	H_2	H_3	H_4	H_5	
1	21.37514	40.25279	15.36849	22.22992	26.28214	21.05198	40.03974	15.20978	22.04453	26.03914	3.848
2	20.74688	39.9778	14.95244	21.97061	25.94238	21.31232	40.22529	15.32689	22.20399	26.24817	3.727
3	21.47224	40.06028	15.344	22.18321	26.0622	21.10213	40.10979	15.17882	22.10856	26.10726	1.998
4	20.91048	40.10609	15.19133	21.95731	26.01251	21.16933	40.09787	15.21861	22.11085	26.08727	1.832
5	21.5028	40.04228	15.14889	22.16699	26.09236	21.16385	40.09355	15.20755	22.09343	26.07656	2.429
6	21.40367	40.23477	15.4103	22.36709	26.26783	21.22173	40.10254	15.22196	22.12816	26.09727	3.909
7	21.53276	40.23035	15.32421	22.16859	26.27788	21.26496	40.12414	15.24474	22.14813	26.1267	6.781
8	21.14701	40.1041	14.79022	21.71361	26.26096	21.26464	40.12669	15.20269	22.10555	26.14661	10.135
9	21.44861	40.16828	15.59713	22.16583	26.17052	21.2801	40.13028	15.23182	22.10178	26.15186	11.516
10	20.32414	39.9469	14.69693	22.21045	25.84648	21.18637	40.11237	15.18239	22.11336	26.12153	8.933
11	21.47476	40.33704	16.09065	22.44353	26.34193	21.21521	40.13483	15.27322	22.14638	26.14357	10.719
12	21.22772	40.02005	15.25308	22.12308	25.89487	21.2141	40.12152	15.26377	22.14135	26.11689	8.983
13	20.99839	40.07298	15.38705	22.26694	26.12767	21.19231	40.11835	15.27628	22.15421	26.12167	9.271
14	21.24963	40.15326	15.23967	22.36554	26.06599	21.20251	40.12289	15.27006	22.17274	26.11597	10.869
15	21.12245	39.97848	15.3732	21.93203	25.95203	21.19316	40.10758	15.28124	22.14316	26.101	8.131
16	21.82112	40.15864	15.32999	22.24247	26.27943	21.25831	40.11699	15.28305	22.16013	26.12381	13.387
17	20.94721	40.07874	15.3307	22.26007	26.03262	21.2061	40.11161	15.28606	22.16704	26.10885	10.957
18	20.73335	40.06492	15.37731	22.12997	26.00501	21.16948	40.10829	15.29335	22.1595	26.10161	9.724
19	21.35438	40.34669	15.51139	22.18066	26.35229	21.20736	40.13406	15.31142	22.16293	26.13097	15.388
20	21.3523	40.17716	15.04093	21.97396	26.14652	21.21489	40.1283	15.2749	22.14319	26.12204	15.37

Figure 5.3: Plots of the memory-type charts of the bimetal dataset.



Although the MEWMA chart generally performed better than the other memory-type control charts to detect moderate-to-large shifts in the mean vector, the MHWMA chart was superior to the other methods when interest lies in detecting a small shift in the mean vector. Furthermore, the HWMA vector elements (in Tables 5.10, and 5.11) give an indication of the direction that the mean has shifted. This indication of the direction of the shift is common among memory-type control chart.

The interpretation of out-of-control signals from multivariate control charts can be quite difficult. For a univariate control chart, an out-of-control state can be easily detected and interpreted, since a univariate chart is associated with only a single variable. However, this is not the case for the multivariate charts.

Because the charts involve a number of correlated variables, the identification and interpretation of any out-of-control signals are not straightforward and has been an interesting topic in SPC literature. We refer the interested reader to Bersimis et al. (2007) for guidance and recommendation on interpreting out-of-control signals in multivariate control charts. In line with Lowry et al. (1992), we recommend monitoring the principal components if these are interpretable. Different researchers, including Jackson (1991), Kourti and Macgregor (1996), and Maravelakis et al. (2002), among others, have proposed various principal-component methods to aid interpretation of out-of-control signals. For example, an MHWMA chart based on the first k principal components or the joint univariate control charts with standard or Bonferroni control limits across the p variables can be plotted.

5.7 Conclusion

In this chapter, a new multivariate chart, namely, multivariate homogeneously weighted moving average (MHWMA) control chart, is proposed for the monitoring of process mean vector. The performance of the chart is evaluated and compared with multivariate χ^2 , MEWMA, MCI and MCUSUM charts considering a variety of charting parameters. The run length comparison revealed that the proposed MHWMA chart is superior to the compared charts, particularly for the detection of small shifts in the process mean vector.

Chapter 6

A new distribution-free multivariate control chart for ecological applications

This chapter is about to be submitted for publication. Also, the chapter will be implemented in a statistical package.

Environmental monitoring programmes have been established across the globe to monitor whole communities of species through time. It is vitally important that such programmes be empowered to detect significant changes in sustainable ecosystems quickly, so that deeper investigations and remedial restoration measures, if necessary, can be implemented directly. In industrial monitoring, a commonly used tool for an analogous purpose is the control chart. The construction of the traditional control chart is equivalent to the plotting of the acceptance regions from a sequence of parametric models over time. A system is considered to be out of control when an observed value falls outside the control-chart limits obtained when there is no impact on the system. When this occurs, it suggests that the system has changed, and appropriate investigations should be initiated. Analogous to the industrial monitoring, in this work, we

propose a multivariate control chart for monitoring ecological data sets. The monitoring aims to discover when (and where) an impact may have occurred in the ecosystem. Our proposed charting method is based on a change point method which does not require prior knowledge of the ecosystem’s behavior before the monitoring can begin. Also, the method does not require any parametric model assumptions, as it can be based on any dissimilarity measure of choice. A permutation procedure is employed to obtain the control-chart limits for a suitable distance-based model through time. We provide examples to show the applications of the method in detecting shifts in ecological community structure. Our simulation results and results of the application of proposed method show that the method detects shifts in the ecosystem earlier than the already available method based on the distance to centroid of the multivariate ecological dataset.

6.1 Introduction

Baseline ecological monitoring has become essential for ecological and environmental sciences, due to the increase in anthropogenic pressure on ecosystems (Anderson and Thompson, 2004). Such monitoring is particularly important where there is concern about potential anthropogenic impacts on ecosystems. Baseline monitoring provides the necessary background information for assessing changes in the ecosystem, and enables ecological impact assessments. As commented by Anderson and Thompson (2004), “one might argue that the role of monitoring (as the name suggests) is to provide a “signal” or “alarm bell” to the presence of an impact, if, where, and when it does occur. That is, we should hope that a reasonable monitoring program would provide us with a way of assessing, at any particular time, whether a measurement we observe is unusual, given what we would expect from our observations of the naturally variable system up until that time.”

The monitoring of changes in production processes is a similar problem with many real-life applications

such as industrial and manufacturing applications. The demands for high product quality have led to the development of many shifts detection methods in the industrial and manufacturing applications (Cui et al., 2008). Statistical process control (SPC) is a fundamental methodology consisting of many standard methods that have been proven useful in quality and productivity improvement of products and processes. Among these methods, the control chart is the most popular and sophisticated tool for tracking and keeping processes in control by monitoring essential quality characteristics of interest (Yen and Shiau, 2010). The construction of the traditional control chart is equivalent to plotting the acceptance regions of a sequence of parametric models over time. A process is considered to be out of control when an observed value or future obtained value of the process falls outside the control-chart limits obtained when there is no changes in the process. When this occurs, it suggests that the process may have been affected by some factors, and thus, appropriate investigation of the process should be initiated.

Control charts as monitoring tools have proven very useful for both the univariate and multivariate independent normal or “approximately” normal distributed data (Montgomery, 2009). The statistical properties of the traditional control charts are accurate only if this assumption is satisfied. However, in some cases, the underlying system is not normal, and as a result, the traditional charts can be very affected in such situations (Chakraborti et al., 2001). Also, tests for normality may be impossible to achieve on a routine basis. “I cannot believe that there are tests for multivariate normality with sufficient power for practical sample sizes that I would even bother to use them; distribution-free multivariate SPC is what we need” (Coleman as cited in Montgomery and Woodall (1997)). The necessity of this assumption has almost restricted the applications of the traditional control charts to industrial process monitoring (Tuerhong and Bum Kim, 2014).

To address this problem, many distribution-free control charts have been proposed. See Chakraborti et al. (2001) for a detailed review of works on distribution-free univariate control charts. Also, several nonparametric multivariate principal components analysis (PCA) based control charts have been pro-

posed by different authors, including Phaladiganon et al. (2013), Martin and Morris (1996), Babamoradi et al. (2016). The PCA based control charts have the ability to handle large numbers of highly correlated variables, measurement errors and missing data (Jackson, 1991). They can improve the ability to detect faults early and detect changes in the process variables (MacGregor et al., 1994; Kourti, 2005), and have found great application in monitoring chemical processes. Moreover, several multivariate non-parametric control charts have been proposed by different authors, including Tsung and To (2012); Zou et al. (2012); Phaladiganon et al. (2013); Qiu (2008). These nonparametric charts are designed for processes, characterized by continuous quality characteristics, and are suitable only for industrial and manufacturing process or product monitoring (Tuerhong and Bum Kim, 2014). They cannot handle multivariate observations of species abundance data; like those obtained for monitoring ecological communities (Anderson and Thompson, 2004).

To make control charts more useful for ecological monitoring, what is needed is a robust non-parametric method which is capable of dealing with multivariate observations of species abundance. Few researchers have proposed different control charts tools for monitoring the ecosystems. Schipper et al. (1997) proposed the use of a sequential probability ratio test (SPRT). Also, Pettersson (1998) investigated the use of statistical surveillance in the form of Hotelling's T^2 and the Shannon-Wiener index for monitoring biodiversity. However, their approaches cannot handle multivariate species abundance data, and also, they required distributional assumptions of the datasets.

Recently, Anderson and Thompson (2004) proposed a distance-based multivariate control chart tool for ecological monitoring. Their method was designed to detect impacts at several spatial scales in hierarchical designs. Moreover, the method allowed any dissimilarity measure of choice, as it does not require any distributional assumptions, since, the control-limits of the chart were obtained using a bootstrap approach. However, because their proposed charting methodology is based on distances to centroid, it failed to account for anisotropic variation in the multivariate data (see Figure 6.1). That is, their control chart

method considered equidistant points from the centroid of the baseline data to be equivalent, even if one point was unusual with respect to the baseline distribution.

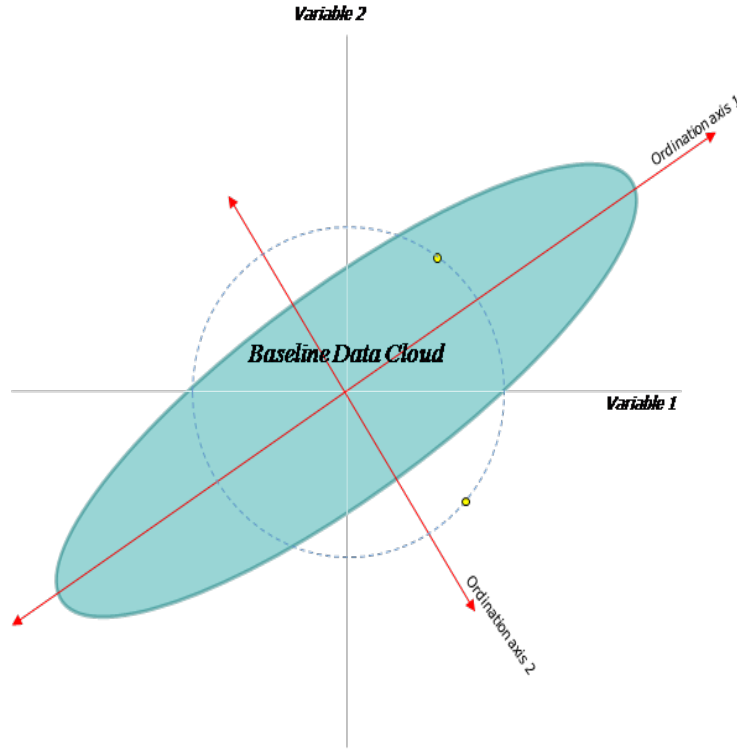


Figure 6.1: Two observations may be equidistant from the centroid - this does not mean they will both lie within (or outside) the baseline data cloud.

Hence, in this work, we propose a more useful multivariate control chart tool for ecological monitoring. Here, the proposed tool is based on a change point method where the present observed ecological data point is evaluated based on all of the previous data points. The method can take either multivariate dataset, or, the dissimilarities between every possible pair of samples as input. When dissimilarities between every possible pair of the samples are available, the method first represents these sample distances as points in coordinate space. Then, these coordinates are used to monitor and identify, as quickly as possible, changes in the mean vector of the multivariate ecological dataset. Importantly, the proposed method can handle anisotropic variability by accounting for the directionality of variation. We use a permutation procedure to obtain the proposed chart limits.

The remainder of the chapter is organized as follows. We present relevant background methodology in Section 6.1.1. Section 6.2 describes the design of the proposed charting tools under both parametric and non-parametric cases, followed by a review of a useful tool for ecological monitoring in Section 6.3. Simulation studies are provided in Section 6.4, followed by ecological application in Section 6.5, and conclusion and discussion is given in 6.6

6.1.1 Background methodologies

6.1.2 The multivariate chi-square chart

Given \mathbf{y}_t ($t = 1, 2, \dots, n$) as a data matrix of a sequence of observations that is distributed as a multivariate normal distribution with mean vector $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Sigma}$, i.e., $\mathbf{y}_t \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. The leading diagonal of $\boldsymbol{\Sigma}$ contains the variances of the columns of \mathbf{y} 's and the off-diagonal elements are the covariances. To detect change in the mean vector, $\boldsymbol{\mu}$, of the multivariate data matrix \mathbf{y}_t , Hotelling (1947) proposed a multivariate control chart tool which represents the weighted Mahalanobis distance of the sample points from the center of the cloud. The tool is known as the multivariate χ^2 control chart, and is given as:

$$\chi^2_t = (\mathbf{y}_t - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{y}_t - \boldsymbol{\mu}) \quad (6.1)$$

The chart signals whenever the χ^2_t values obtained from the observation at time t is greater than the chart's constant control limit $\chi^2_{p,\alpha}$. The quantity $\chi^2_{p,\alpha}$ is the α^{th} upper percentage point of the chi-square distribution and p is the number of columns of the data matrix \mathbf{y}_t .

The chart has been found useful in a range of scientific and technological application domains, including health-related monitoring, quality improvements, ecological monitoring, spatiotemporal surveillance, and profile monitoring. Hotelling (1947) employed the chart to bombsight data during World War II. Since then, several papers dealing with control procedures for several related variables have appeared in

literature and include Anderson and Thompson (2004), Crosier (1988), Pignatiello and Runger (1990), Roberts (1959), Jackson (1991), Woodall and Ncube (1985), Golosnoy et al. (2009), Zhang and Chang (2008), Phaladiganon et al. (2013), Ngai and Zhang (2001), and Hawkins (1991). We refer to Bersimis et al. (2007) for a recent review in multivariate control charting procedures.

The multivariate χ^2 chart is only applicable when the parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are known. In practice, the parameters are generally unknown and they need to be estimated from historical in-control behaviour of the system. In this case, the chart is applied in a two-phase scheme, called, Phase I and Phase II. The Phase I scheme is used to model the in-control performance of the process based on the historical sample dataset, whereas in Phase II, new observations (not used in model building) are checked against the chart's performance obtained from Phase I. A direct approach to estimate the parameters is to obtain their empirical estimates. However, the empirical estimate of the covariance matrix leads to a chart that may not be efficient in detecting a shift in the mean vector, because the parameter may be poorly estimated (Sullivan and Woodall, 1996). To enhance the performance of the chart under estimated parameters, different estimates of the covariance matrix have been suggested. For example, Sullivan and Woodall (1996) proposed a method of estimation that uses the differences between successive observations. Also, Holmes and Mergen (1993) proposed the use of a mean square successive difference approach for estimating $\boldsymbol{\Sigma}_0$.

When the parameters are unknown and estimated from Phase I, to avoid estimation bias during the model building in Phase I and make chart sensitivity to change in the mean vector, the chart parameters $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$, need to be estimated from a large in-control reference sample. However, in some practical applications, information about the Phase I behaviour of the system may not be available, or it may not be possible to wait and gather a reasonably large historical in-control sample about the process before commencing the monitoring process. To solve this problem, Quesenberry (1991, 1997) proposed methods that update the parameter estimators of the mean vector and covariance matrix with new observations

and simultaneously monitor the process to determine whether the new observation is in control or not. A variant of these methods based on sample distances is employed in the multivariate control charts tools for ecological and environmental monitoring by Anderson and Thompson (2004).

6.2 Description of the proposed method

6.2.1 Parametric case

To detect change in the mean vector $\boldsymbol{\mu}$ of the multivariate data matrix \mathbf{Y} , following Anderson and Thompson (2004), at any particular time t , we assess the extent to which the observation, \mathbf{y}_t , is unusual, given previous values up to (and including) time $(t - 1)$. That is, we define a vector, \mathbf{z}_t , as the weighted deviation of the observation t from the average of all the previous observations. This is given as:

$$\mathbf{z}_t = (\mathbf{y}_t - \bar{\mathbf{y}}_{t-1}) \sqrt{\frac{t-1}{t}} \quad (6.2)$$

where $\bar{\mathbf{y}}_{t-1} = \frac{1}{t-1} \sum_{j=1}^{t-1} \mathbf{y}_j$, and $\mathbf{y}_0 = \mathbf{0}$.

We calculate the estimate of the mean vector of \mathbf{Y} at each value of t using the updated formulas in Quesenberry (1991, 1997), given as:

$$\bar{\mathbf{y}}_t = \frac{1}{t} ((t-1)\bar{\mathbf{y}}_{t-1} + \mathbf{y}_t), \quad (6.3)$$

where $t = 2, 3, 4, \dots$ for the estimated mean vector. Hence, the $\bar{\mathbf{y}}_{t-1}$ in Equation(6.2) are calculated using the formula given in Equation (6.3). We assume that the system is “in control”, and that the \mathbf{y}_t are stochastic with a common expected value, estimated by $\bar{\mathbf{y}}_{t-1}$. Hence, under this null hypothesis, the \mathbf{z}_t are stochastic deviations with an expected value of zero, and are independent and identically distributed as a multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $\mathbf{z}_t \sim N(\mathbf{0}, \boldsymbol{\Sigma})$.

An unusually large value of \mathbf{z}_t signals that the system is “out of control” at time t .

When the process is in-control, the vector \mathbf{z}_t can be transformed to give:

$$Q'(\mathbf{z}_t) = \Phi^{-1} \left[F_{p, t-p-1} \left(\frac{(t-p-1)}{p(t-2)} \right) \times \left(\mathbf{z}_t^T \mathbf{S}_{\mathbf{z}_{t-1}}^{-1} \mathbf{z}_t \right) \right] \quad (6.4)$$

where p is the number of variables in \mathbf{Y} , $t = p+2, p+3, \dots$, $F_{p, t-p-1}$ is the cumulative distribution function (cdf) of F distribution with p and $t-p-1$ degrees of freedom, and Φ^{-1} is the inverse of the cdf of the standard normal distribution (Quesenberry, 1991, 1997). Using the transformation in Equation (6.4), the in-control transformed chart statistic $Q'(\mathbf{z}_t)$ follows a standard normal distribution (i.e., $Q'(\mathbf{z}_t) \sim N(\mathbf{0}, \mathbf{I})$). $\mathbf{S}_{\mathbf{z}_t}$ in Equation (6.4) is the sample (or empirical) covariance matrix of \mathbf{z}_t at time t , and is obtained as follows:

$$\mathbf{S}_{\mathbf{z}_t} = \frac{1}{t-1} \mathbf{H}_t \quad (6.5)$$

where \mathbf{H}_t given by $\mathbf{H}_t = \mathbf{H}_{t-1} + \mathbf{z}_t \mathbf{z}_t^T$, $t = 3, 4, \dots$, and $\mathbf{H}_0 = \mathbf{0}_{k \times k}$ (Quesenberry, 1991, 1997).

The squared standardized (generalized) distance of \mathbf{z}_t to its expected value (i.e., the Hotelling's T^2 statistic of \mathbf{z}_t), is given as:

$$T_t^2 = \mathbf{z}_t^T \mathbf{S}_{\mathbf{z}_{t-1}}^{-1} \mathbf{z}_t \quad (6.6)$$

Because the T_t^2 statistic in Equation (6.6) is based on updating formulas for the estimators of the mean vector and covariance matrix, the T_t^2 statistic is not distributed as a $F_{p, t-p-1} \left(\frac{(t-p-1)}{p(t-2)} \right)$. Hence, we transformed the T_t^2 statistic to

$$Q'(T_t^2) = (\chi_1^2)^{-1} \left[F_{p, t-p-1} \left(\frac{(t-p-1)}{p(t-2)} \right) \times T_t^2 \right] \quad (6.7)$$

where $(\chi_1^2)^{-1}$ is the inverse of the cdf of a chi-square distribution with one degree of freedom. When the process is in control, then, the value of the $Q'(T_t^2)$ should be less than the upper control limit $\chi_{\alpha,1}^2$, where

$\chi^2_{\alpha,1}$ is the upper $(\alpha)th$ percentage point of the chi-square distribution with one degree of freedom. The chart gives out-of-control signal whenever the $Q'(T^2_t)$ value is greater than the chart upper control limit (i.e., $\chi^2_{\alpha,1}$). Here, we recommend plotting the transformed statistic $Q'(T^2_t)$ in Equation (6.7), when the data matrix \mathbf{y}_t is distributed as a multivariate normal with unknown mean vector and covariance matrix. We refer to the proposed chart based on the $Q'(T^2_t)$ as a change point monitoring tool based on a chi square distribution with one degree of freedom.

6.2.2 Non-parametric case

Retrospective analysis

Here, \mathbf{Y}_t is a multivariate observation obtained from discrete sequential time points, $t = 1, 2, \dots$, and consists of counts of abundance \mathbf{y}_{tk} on each of $k = 1, 2, \dots, p$ species or taxa. In this case, we first represent the species abundance data matrix, \mathbf{y}_t , as points in multivariate ordination space $\mathbf{Q}_{tk'}$ (here, k' is the dimension of $\mathbf{Q}_{tk'}$), using appropriate ordination tool. These tools are based on pattern extraction (Legendre and Legendre, 1988), and have found significant applications in Ecology. They used the relationships among the original variables to build new variables which describe and carry a dominant part of the global pattern in the original variables (McCune, Bruce and Grace, James B and Urban, 2002).

An important ordination method is the metric multidimensional scaling (MDS); a method initially proposed by Kruskal (1964a,b). For a given configuration, the degree to which the MDS inter-points distances is a function of the experimentally observed dissimilarity is expressed by a measured called stress (S). A large value of S symbolizes a poor fit, whereas a small stress value means a good fitting solution (Legendre and Legendre, 1988). Also, Gower (1966) described a method to represent a multivariate sample of size n , as points in a Euclidean space. This method, known as principal coordinate analysis (PCO), takes an input matrix giving dissimilarities between n pairs of samples. We refer the interested reader to Legendre and Legendre (1988); Kruskal (1964a); Anderson and Willis (2003); McCune, Bruce and Grace,

James B and Urban (2002); Frontier (1976); Bajorski (2011); Franklin et al. (1995); Paul and Anderson (2013); Jackson (1993) for literature on a number of important ordinations tools commonly applied in practice, and on the determination of the number of nontrivial ordination axes, k' , to be retained in $\mathbf{Q}_{tk'}$. Using the $\mathbf{Q}_{tk'}$, we define the statistic \mathbf{z}_t given as:

$$\mathbf{z}_t = (\mathbf{q}_t - \bar{\mathbf{q}}_{t-1}) * \sqrt{\frac{t-1}{t}} \quad (6.8)$$

where \mathbf{q}_t is the t th row of the matrix $\mathbf{Q}_{tk'}$, $\bar{\mathbf{q}}_{t-1} = \frac{1}{t-1} \sum_{j=1}^{t-1} \bar{\mathbf{q}}_j$ and $\mathbf{q}_0 = \mathbf{0}$. The monitoring statistic is given as:

$$T_t^2 = \mathbf{z}_t^T \hat{\Sigma}_{\mathbf{z}_{t-1}}^{-1} \mathbf{z}_t \quad (6.9)$$

In some cases, the number columns of the matrix $\mathbf{Q}_{tk'}$ (*i.e.*, k'), would be close to the number of rows (say, n). This makes the usage of the empirical estimate (say \mathbf{M}_t) of the covariance matrix of $\mathbf{Q}_{tk'}$ unsuitable, for relatively small t . This is because, the inverse of empirical estimate of the covariance matrix would not exist unless a reasonable number of samples are gathered before the monitoring schemes begins. Hence, the matrix $\hat{\Sigma}_{\mathbf{z}_t}$ in Equation (6.9) is given as:

$$\hat{\Sigma}_{\mathbf{z}_t} = \frac{1}{t-1} \mathbf{W}_t \quad (6.10)$$

where $\mathbf{W}_t = \mathbf{W}_{t-1} + \mathbf{z}_t \mathbf{z}_t^T$ and $\mathbf{W}_0 = \mathbf{0}_{k' \times k'}$

The matrix \mathbf{W}_t (in Equation (6.10)) is a shrinkage estimate of the covariance matrix and its usage particularly helps to commence the monitoring scheme for relatively small samples. It is a weighted average of the empirical estimate \mathbf{M}_t and a target matrix $\mathbf{T} = (t_{ij})_{1 \leq i, j \leq r}$; is given as:

$$\mathbf{W}_t = \lambda \mathbf{T} + (1 - \lambda) \mathbf{M}_t, \quad (6.11)$$

where $\lambda \in [0, 1]$ denotes the shrinkage intensity: $\lambda = 0$ implies $\mathbf{W}_t = \mathbf{M}_t$, and $\lambda = 1$ gives $\mathbf{W}_t = \mathbf{T}$. The estimator is well conditioned for small sample, and does not make any distributional assumption about the underlying distribution of the data. Its performance advantages are, therefore, not restricted to multivariate normal data (Ledoit and Wolf, 2004a, 2003; Schäfer and Strimmer, 2005; Ullah et al., 2017; Adegoke et al., 2018b). The target matrix, \mathbf{T} , used is the one that shrinks the diagonal elements (i.e., the sample variances) of the empirical estimate of the covariance matrix towards their median and shrinks the off-diagonal entries to zero. Other targets (see Ledoit and Wolf (2004a); Schäfer and Strimmer (2005)) like shrinking the diagonal against zero or towards mean can also be considered. Opgen-rhein and Strimmer (2007) showed that these two alternatives turned out to be either less efficient (zero target) or less robust (mean target) than shrinking towards the median. The shrinkage covariance matrix has the effect of reducing the larger eigenvalues and increasing the smaller ones, thereby counteracting biases inherent in sample-based estimation of eigenvalues (Friedman, 1989; Opgen-rhein and Strimmer, 2007).

A key question to the usage of the shrinkage estimate in Equation (6.11) is the determination of the optimal shrinkage intensity parameter (λ^*). Several studies have proposed different methods to estimate this parameter. For example, Ledoit and Wolf (2004a); Schäfer and Strimmer (2005); Opgen-rhein and Strimmer (2007) gave analytic approach for determining the optimal shrinkage intensity parameter. Friedman (1989) proposed a computationally very intensive approach to estimate λ^* using cross-validation. Morris (1983), and Greenland (2000) estimated λ^* in an empirical Bayes context. We follow the approach in Ledoit and Wolf (2004a); Opgen-rhein and Strimmer (2007), and derive the estimated optimal intensity parameter analytically. Here, the optimal value of the shrinkage intensity, λ^* , is obtained by minimizing the risk function:

$$R(\lambda) = \|\mathbf{W}_t - \mathbf{\Sigma}\|_F^2 \quad (6.12)$$

where $\mathbf{\Sigma}$ is the true covariance matrix, and $\|\cdot\|_F^2$ is the squared Frobenius norm, which is a quadratic

form associated with the inner product. The optimal intensity parameter in this case is given as:

$$\hat{\lambda}^* = \frac{\sum_{t=1}^p \hat{Var}(m_{tt})}{\sum_{t=1}^p (m_{tt} - \text{median}(m))^2} \quad (6.13)$$

where $\text{median}(m)$ is the median of the sample variances of \mathbf{M}_t (Opgen-rhein and Strimmer, 2007; Ullah et al., 2017). The T_t^2 statistic in Equation (6.9) (obtained based on the shrinkage estimate of the covariance) is defined for $t = 4, 5, \dots, n$, otherwise, we recommend accumulating suitable samples before commencing the monitoring scheme.

Control limits

Here, our aim is to compare the realization of the ecosystem at the time t against the system behaviour prior to the time t . Under the null hypothesis (i.e., if the system is “in-control”), the distribution of the y_t space can be modeled as stochastic process through time. We propose a method to obtain the chart limit at any particular time t , using a permutation procedure. The permutation tests are very useful to obtain reliable statistical inferences without making strong distributional assumptions (Hessainia et al., 2013), except for the exchangeability of the observational data (Anderson, 2001). We refer interested reader to Anderson and Ter Braak (2003); Anderson (2001) for different permutation strategies commonly used in practice.

In addition to the null hypothesis, we obtain the limit by assuming that any of the observations (in $\mathbf{Q}_{tk'}$) prior to time t , can occur at time $t - 1$. Under this assumption, the limit is obtained by permuting the rows of $\mathbf{Q}_{tk'}$ up to and including the $(t - 1)th$ row, and calculating T_{t-1}^2 . The T_{t-1}^2 is the value of the proposed charting statistic (given in Equation 6.9) for the last permuted observation (at time $t - 1$). Repeat this permutation procedure many times to get the distribution values of 95% percentile of T_{t-1}^2 . We refer to this 95% percentile of the distribution of T_{t-1}^2 as $\delta^{T_{t-1}^2}$. The $\delta^{T_{t-1}^2}$ is used as the upper control

limit of the chart for monitoring the system at time t . In the derivation of the chart' limit, because the T_t^2 is defined for $t = 4, 6, \dots, n$, the limits vary with time, are defined only for $t \geq 5$, and hence, the monitoring scheme begins at $t = 5$.

6.2.3 Progressive analysis

When the proposed chart is based on the ordination axis $\mathbf{Q}_{tk'}$, an important question with the calculation of the proposed charts given in Section 6.2.2, is how to interpret the control chart's behaviour at time t given that the system is already out of control at time $t-1$. That is, given that the behaviour of the system at time $t-1$, calculated from $\mathbf{Q}_{tk'}$, gives an out of control signal, then, it is expected that the observation at time $t-1$ should not be included in the calculation of the chart's limit at time t . Hence, instead of calculating the proposed chart statistic and limits on the matrix, $\mathbf{Q}_{tk'}$, obtained from all the rows of the data matrix \mathbf{y}_t , we recommend calculating separate matrices $\mathbf{Q}_{t^*k'}$ for each time $t^* = 5, 6, 7, \dots, n$. Here, the matrix $\mathbf{Q}_{t^*k'}$ used for monitoring the observation at time t^* , is obtained from only the first t^* samples (or rows) of \mathbf{y}_t .

As a result, the proposed chart statistic for monitoring the observation at time, t^* , is calculated only from the last observation in $\mathbf{Q}_{t^*k'}$, and is given as:

$$\mathbf{z}_{t^*} = (\mathbf{q}_{t^*} - \bar{\mathbf{q}}_{t^*-1}) \sqrt{\frac{t^* - 1}{t^*}} \quad (6.14)$$

where \mathbf{q}_{t^*} is the last row of $\mathbf{Q}_{t^*k'}$, $\bar{\mathbf{q}}_{t^*-1} = \frac{1}{t^* - 1} \sum_{j=1}^{t^*-1} \bar{\mathbf{q}}_j$ and $\mathbf{q}_0 = \mathbf{0}$. In this case, the monitoring statistic (for time t^*) is standardized by the dimension of matrix $\mathbf{Q}_{t^*k'}$ (at time t^*), and is given as:

$$T_{t^*}^2 = \frac{1}{k'} \mathbf{z}_{t^*}^T \hat{\Sigma}_{\mathbf{z}_{t^*-1}}^{-1} \mathbf{z}_{t^*} \quad (6.15)$$

where k' is the dimension of $\mathbf{Q}_{t^*k'}$ at time t^* , $\hat{\Sigma}_{\mathbf{z}_{t^*-1}} = \frac{1}{t^* - 1} \mathbf{W}_{t^*}$, $\mathbf{W}_{t^*} = \mathbf{W}_{t^*-1} + \mathbf{z}_{t^*} \mathbf{z}_{t^*}^T$ and $\mathbf{W}_0 = \mathbf{0}_{k' \times k'}$.

Also, the control limit for monitoring the behaviour at time t^* is calculated from the $\mathbf{Q}_{t^*k'}$ (and $T_{t^*}^2$ given in Equation (6.15)), using the control chart limit procedure explained in Section 6.2.2. We refer to the proposed chart given in this Section as the change point chart based on the permutation limit (or chart based on the permutation limit).

In all of the analysis in this study, at any particular time t (or t^*), the dimension of the matrix $\mathbf{Q}_{tk'}$ (or $\mathbf{Q}_{t^*k'}$) was obtained such that the linear correlation between the distances of the matrix $\mathbf{Q}_{tk'}$ (or $\mathbf{Q}_{t^*k'}$) and the original data matrix is 0.999.

6.3 Distance to centroid approach

At any particular t , the distance to the centroid of all previous observations (d_t) method by Anderson and Thompson (2004) is given as:

$$d_t = \sqrt{\frac{t}{(t-1)}(SS_t - SS_{t-1})} \quad (6.16)$$

where, $SS_t = \frac{1}{t} \sum_{i=1}^{t-1} \sum_{j=i+1}^t D_{ij}^2$, and $SS_{t-1} = \frac{1}{(t-1)} \sum_{i=1}^{t-2} \sum_{j=i+1}^{t-1} D_{ij}^2$ are sum of squared interpoints dissimilarities among all points up to time t divided by t and the sum of squared interpoint dissimilarities among all points up to time $(t-1)$, divided by $(t-1)$, respectively. The control chart's upper confidence limit for d_t was obtained using a bootstrapping procedure. The upper bound of d_t (in Equation (6.16)) was formulated to deal with the situations where there are more than one sites by using the average of the 95th percentile of the empirical distribution of d_t across the entire set of sites in the spatial array (Anderson and Thompson, 2004). As the d_t statistic was based on distances from centroid calculated from all previous observations, the confidence bound of d_t change through time. Detailed information on the derivations of the upper bound of d_t is given in Anderson and Thompson (2004).

Here, we give the following steps for obtaining the upper bound of d_t at time t , when there is only

one site:

- 1 Take \mathbf{y}^* as a random sample point from $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{t-1}$.
- 2 Take a bootstrap sample with replacement from $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{t-1}$, call this matrix $\mathbf{y}^{[b]}$. Note, here, the random sample \mathbf{y}^* is excluded from the matrix $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_{t-1}$ from which $\mathbf{y}^{[b]}$ was obtained.
- 3 Obtain distance to centroid of \mathbf{y}^* from $\mathbf{y}^{[b]}$, call this, $d_t^{[b]}$.
- 4 Repeat steps 1 to 3 many times, and obtain 95% quantile of the empirical distribution of $d_t^{[b]}$.

6.4 Performance of the new method and results

We first examined the performance of the methods using two different simulation studies. The first was based on data drawn from multivariate normal distribution while the second simulation involves a simulated ecological dataset. For the multivariate normal simulated dataset, the performance of the proposed charts: the chart based on the chi-square distribution with one degree of freedom (Section 6.2.1) and the chart based on the permutation limit (Section 6.2.2), are compared with both the control charts based on distance to the centroid of all previous observations (d_t) by Anderson and Thompson (2004) and the multivariate chi-square by Hotelling (1947) explained in Sections 6.3 and 6.1.2, respectively. For the ecological simulated data set, the performance of the proposed chart based on the progressive analysis (Section 6.2.3), is only compared with the distance to the centroid of all previous observations (d_t) charting tool. The proposed chart and d_t statistics were obtained on the basis of the Euclidean distance measure and Bray-Curtis dissimilarity measure for the normally and ecological simulated datasets, respectively.

In both simulations, n_1 (out of $n = n_1 + n_2$) observations were drawn from in-control distribution, and the remaining $n_2 = 1$ observation was drawn from out-of-control distribution. We considered $n_1 = 5, 7, 10, 15, 20, 25, 30, 40$, or 50 in our simulations. In both cases, the charts were designed such that the

false alarm probability was fixed at 0.05. For each of the control chart under investigation, the false alarm rate and power were calculated, respectively, as the percentage of the l (number of simulations) calculated plotting statistic for in-control observations that exceeded the control limit of the chart and the percentage of the l calculated plotting statistic values for out-of-control observations that exceeded the control limit of the chart. The false alarm rate and power were calculated from $l = 10,000$ simulations.

6.4.1 Multivariate normal simulation description and results

We drew the n observations from p -variate ($p = 2$, or 5) normally distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, i.e., $y_t \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Without loss of generality, we used $\boldsymbol{\mu} = \mathbf{0}$ and considered exchangeable covariance matrices $\boldsymbol{\Sigma}$ with variances given as:

$$\sigma_{tk'} = \begin{cases} 1, & \text{if } t = k' \\ b, & \text{if } t \neq k'. \end{cases} \quad (6.17)$$

We examined the performance of the methods under both independent and $AR(1)$ covariance structures, by setting $b = 0$ (i.e., independent covariance structure), and $b = 0.8^{|k-k'|}$ for $1 \leq k, k' \leq p$ (i.e., $AR(1)$ covariance structure), respectively.

To obtain the power from the different charting schemes, we simulated the $n_2 = 1$ observation from the distribution where the mean vector (i.e., $\boldsymbol{\mu}$) was shifted by (size) δ relative to the n_1 dataset. Following, Ullah et al. (2017), we considered three different scenarios for the directions of this shift, along: the first eigenvector, the last eigenvector, and all eigenvectors. In each of these cases, δ was selected so that the simulation was useful in discriminating between the methods. The values of δ used were: $2\sqrt{p\lambda_1}$ along the first eigenvectors, $2\sqrt{p\lambda_p}$ along the last eigenvectors, and $2\sqrt{\lambda_k}$ along all eigenvectors, where λ_k is the k th eigenvalue of $\boldsymbol{\Sigma}$. The results in Figures 6.2 and 6.3 represent the multivariate normally distributed

simulation results (for $p = 2$ and 5) under both the independent and the $AR(1)$ covariance structures, respectively.

As shown in Figure 6.2, the proposed chart method based on the permutation limit and the distance to the centroid of all previous observations (d_t) charting tool gave higher false alarm probabilities than expected for relatively small values of n_1 . In particular, when $p = 2$, the distance to centroid approach gave false alarm probabilities values higher than the expected for all values $n_1 \leq 30$, also, the proposed chart method based on the permutation limit consistently gave the required false alarm probabilities for $n_1 \geq 20$. Hence, the proposed chart based on the permutation limit required fewer sample to achieve the required false alarm probability. However, the chi-square chart and the proposed chart based on the chi square with one degree of freedom consistently gave the required false alarm probabilities for all values of n_1 . Similarly, power probabilities from the chart based on the distance to centroid were invariably higher than the power from the proposed chart based on the permutation for all values of $n_1 \leq 30$, however, for $n_1 > 30$, the chart based on the permutation limit gave higher power. Similar conclusion can be made for the case of $p = 5$.

Figure 6.3 presents the simulation results under the $AR(1)$ structures considered. As shown on the Figure, the power from the distance to centroid depends on the direction of the shift. For both $p = 2$ and 5 , when the direction of the shift is along the first eigenvector, the new method consistently gives higher power than expected (i.e., the results of the chart based on the chi-square with p degrees of freedom (the multivariate chi-square chart)), this is evident in Figure 6.3. Intermediate powers were obtained when the shift is along all the eigenvector. Meanwhile, when the shift is along the last eigenvector, the power probabilities from the method were consistently minimal. The results of the proposed charts based on the chi-square with one degree limit and the permutation limit give almost similar patterns in-terms of their false alarm probabilities and power, for the values of p considered.

6.4.2 Ecological simulated data set and results

Here, we used the multivariate dataset on the putative impact of a shoreline sewage outfall on the structure of molluscan assemblages in a Mediterranean rocky subtidal habitat. The study area is located along the south-western coast of Apulia (Ionian Sea, SE, Italy), and is characterised by wave-exposed calcarenitic rocky plateaus extending from the water surface to about $10m$ depth on fine sand with a gentle-medium slope (Terlizzi et al., 2005b,a). Sampling was undertaken in November 2002 at the outfall location (hereafter indicated as I) and two control or reference locations (C1 and C2, hereafter indicated as Cs). The control locations were chosen at random from a set of eight possible locations separated by at least $2.5km$ and providing comparable environmental conditions to those occurring at the outfall (in terms of slope, wave exposure, type of substrate). They were also chosen to be located on either side of the outfall. At each of the three locations (I and Cs), three sites, separated by $80m - 100m$ were randomly chosen. At I, one site was located immediately adjacent to the point of discharge and the remaining other two were on its right and left, respectively. For each site, assemblages were sampled at a depth of $3m - 4m$ on sloping rocky surfaces. Nine random replicates were collected at each of the three sites within each location, yielding a total of 81 samples. The number of species $p = 151$.

The n_1 and n_2 observations were randomly drawn from the control groups and the impacted group, respectively, using the simulation procedure described in Anderson et al. (2019). We randomly drew the n_1 samples from the control groups, and the $n_1 = 1$ was randomly drawn from the impacted group. Here, the plotting statistic and control limits of our proposed chart were obtained using the progressive analysis of the method given in Section 6.2.3. The matrix $\mathbf{Q}_{t^*k'}$ (MDS axes) used for monitoring the observation at time t was metric MDS axes obtained from the square root transformation of simulated datasets from the control and impacted groups. We provide false alarm probability plot, power plot and a plot of the ratio of the power to the false alarm for both the full dataset and a reduced dataset in Figure 6.4. The reduced data contains a subset of the full dataset with positive covariances.

As shown in the figure (Figure 6.4), the false alarm probabilities from both method were greater than the required 0.05 level. However, the figure shows that the proposed chart required fewer n_1 sample to achieve the required probability level than the method based on the distance to centroid. As expected, the powers obtained from the distance to centroid were higher than those obtained from the proposed chat. Meanwhile, as n_1 becomes larger, the power from the proposed method converges to 1 faster than the method based on the distance to centrid. This is more evident in the power plot results for the reduced dataset. Importantly, the plot of the ratio of the power to the false alarm shows that the proposed method is relatively better than its competitor.

6.5 Ecological example

Here, we used the dataset on fish species sampled annually during the summer by the Department of Fisheries and Oceans (DFO), Canada, for 41 years (1970-2010) in the Scotian Shelf, Northwest Atlantic bottom trawl studies. The studies used a stratified random design to provide unbiased, and independent fishing indices of soil fish abundance and recruitment for the region of Fundy’s Scotian Shelf. The data were mainly benthic fish species, although other species were also routinely captured, including small pelagic fish. All fish caught were classified at species level except for redfish, which were classified as ”unspecified redfish ” because of uncertainties in their taxonomy (Ellingsen et al., 2015).

The Scotian Shelf covers latitudes $42 - 47^\circ N$ and is divided into an eastern region (sub-areas 1 – 3) and a western region (sub-areas 4 – 5, including the Bay of Fundy), corresponding to divisions of the Northwest Atlantic Fisheries Organization (NAFO) with a total area of $172,000km^2$ Figure 6.5. The total area was divided into five sub-areas (Figure 6.5) after the original division into strata by DFO (based on bathymetry, Shackell and Frank (2003)). There are two spatially discrete, reproductively isolated stocks of cod on the Scotian Shelf, one in the eastern half of the shelf and the other in the west. In the early

1990s, only the eastern stock collapsed, and the fishing industry closed there (Frank et al., 2005). Despite being severely depleted, the stock of cod in the western region was never closed to fishing and persisted in a depleted state (Shackell et al., 2010). Cod on the eastern Scotian shelf failed to recover unexpectedly (Frank et al., 2005), although limited recovery was evident after 20 years (Frank et al., 2011). Slow recovery was linked to fundamental changes in the structure of the ecosystem due to excessive rates of use (Frank et al., 2006, 2007, 2011).

Slow recovery has also been linked to predation by gray seals and/or the lack of reproductive success through predator-prey reversals, where forage fish such as herring or capelin consume the eggs, larvae or young-of-the-year cod (Bundy, 2005; Steneck, 2012). While several other commercially exploited groundfish species have also declined in this region (Frank et al., 2005), the benthic community's primary prey has increased, including small benthic and pelagic fish and benthic crustaceans (e.g. snow crab and northern shrimp) (Zwanenburg, 2000; Worm and Myers, 2003; Choi et al., 2004; Frank et al., 2005). Similar ecosystem responses to declining groundfish abundances have also been observed on the western half of the Scotian Shelf, but only minor changes in benthic crustaceans were evident (Shackell et al., 2010).

The data analyses were done on the basis of Bray-Curtis dissimilarity measure (using non-metric MDS), and multivariate control charts based on the distance to centroid of $(t-1)$ observations by Anderson and Thompson (2004), and the proposed chart were provided for the Area 1. The charts from our proposed charting method were based on the progressive analysis given in Section 6.2.3, and the matrix $\mathbf{Q}_{t^*k'}$ used for monitoring the system behaviour at time t^* was MDS axes calculated from the first- t^* observations of the data matrix. At any particular time t^* , the dimension of the matrix $\mathbf{Q}_{t^*k'}$ was obtained such that the linear correlation between the distances of the matrix $\mathbf{Q}_{t^*k'}$ and the original data matrix is 0.999. At any particular time t^* , when the plotting statistic values (or points) are below the chart's control limits, then we assume that the observation at the time t^* is in-control given what were observed prior to that time.

Figures 6.6 presents the results for the Area 1. Addition of a new observation increases dimension of the MDS axes by one unit for the first few years. However, subsequent changes in the dimension of the MDS axes occur on average after every two observations. This indicates that inclusion of an addition observation into the data matrix from which the MDS axes were obtained increases the dimension of the data matrix $Q_{t \times k'}$ used in our proposed charting scheme. For this area, the method based on distance to centroid of $(t - 1)$ observations gives the first "out-of-control" prior to the stock collapse in 1991. In particular, the method shown the first out-of-control point in 1974. However, the proposed chart gives the first "out-of-control" in immediately after the collapse in 1991.

6.6 Conclusion and discussion

In this chapter, we proposed a multivariate control chart tool for ecological monitoring. The proposed chart was based on a change point method where the present observed ecological data point was evaluated based on all of the previous data points. The method was designed to take both the multivariate dataset, and also, the dissimilarities between every possible pairs of samples as input. A permutation procedure was employed to obtain the control-chart limits for a suitable distance-based model through time. Simulation results showed that the method performed better than already available tool used for the same purpose especially when there are structures in the ecological dataset. We provided examples to show the applications of the procedure in detecting real shifts in ecological community structure.

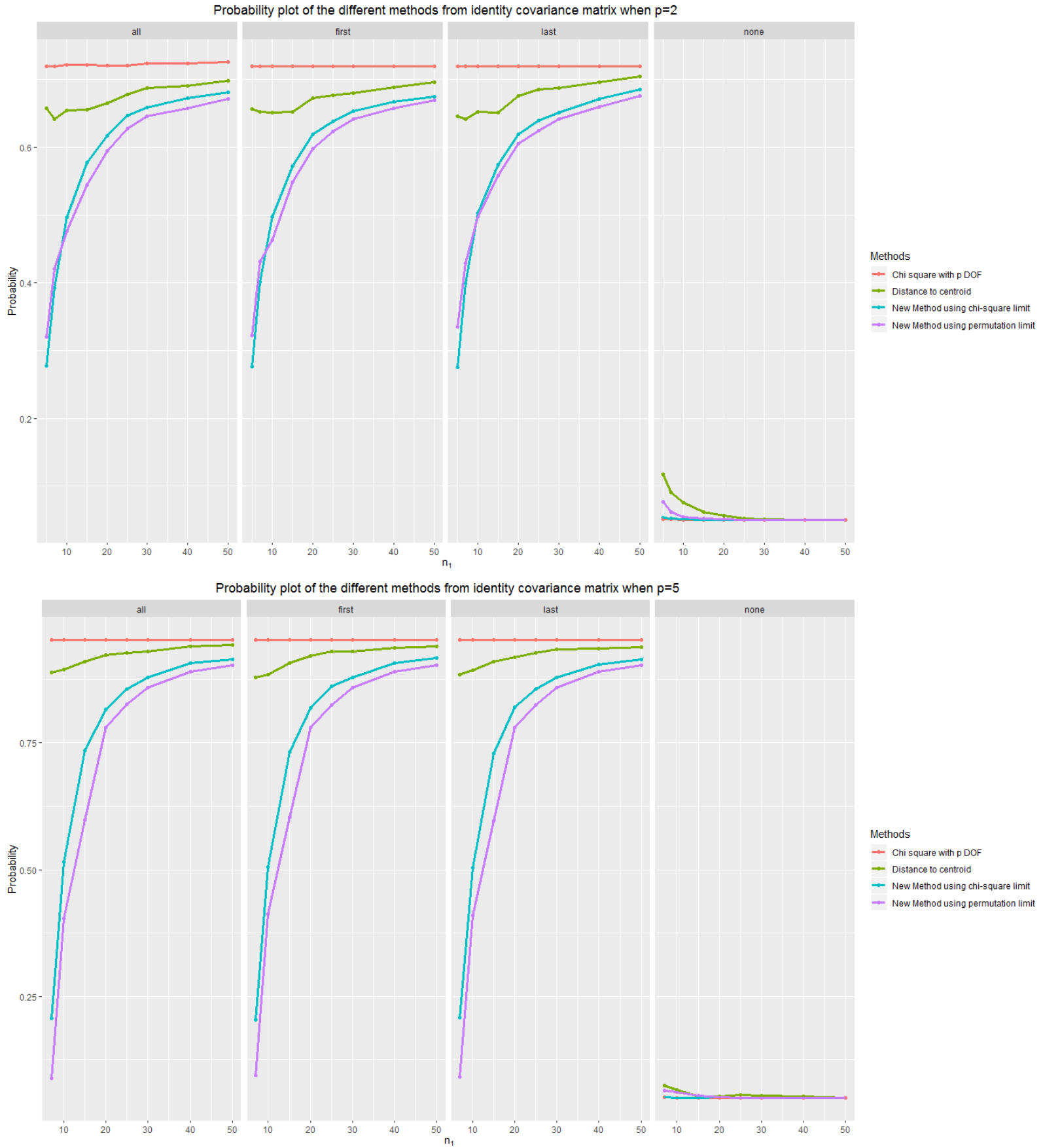


Figure 6.2: Probability plot of the different methods from identity covariance matrix

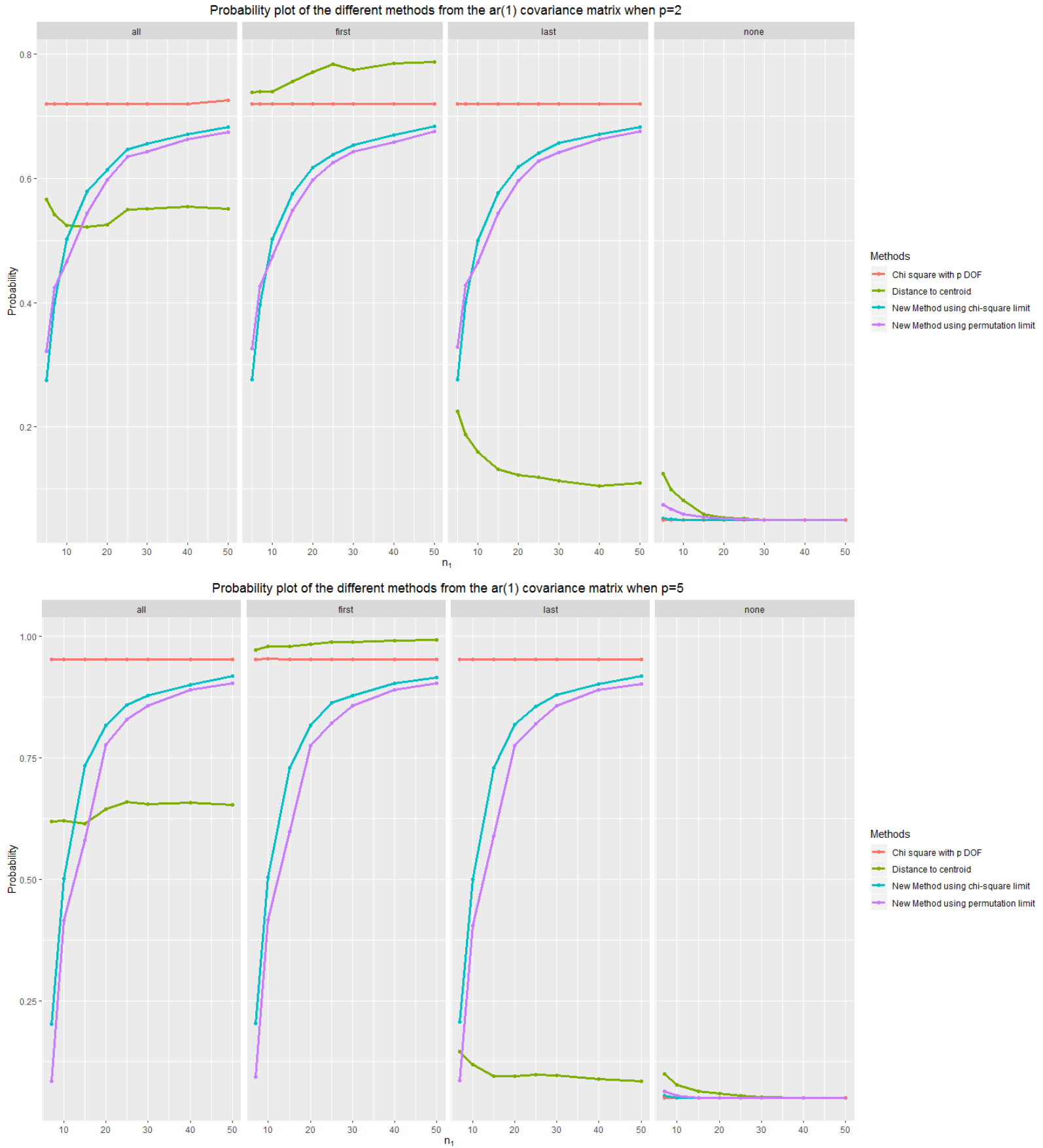


Figure 6.3: Probability plot of the different methods from the ar(1) covariance matrix

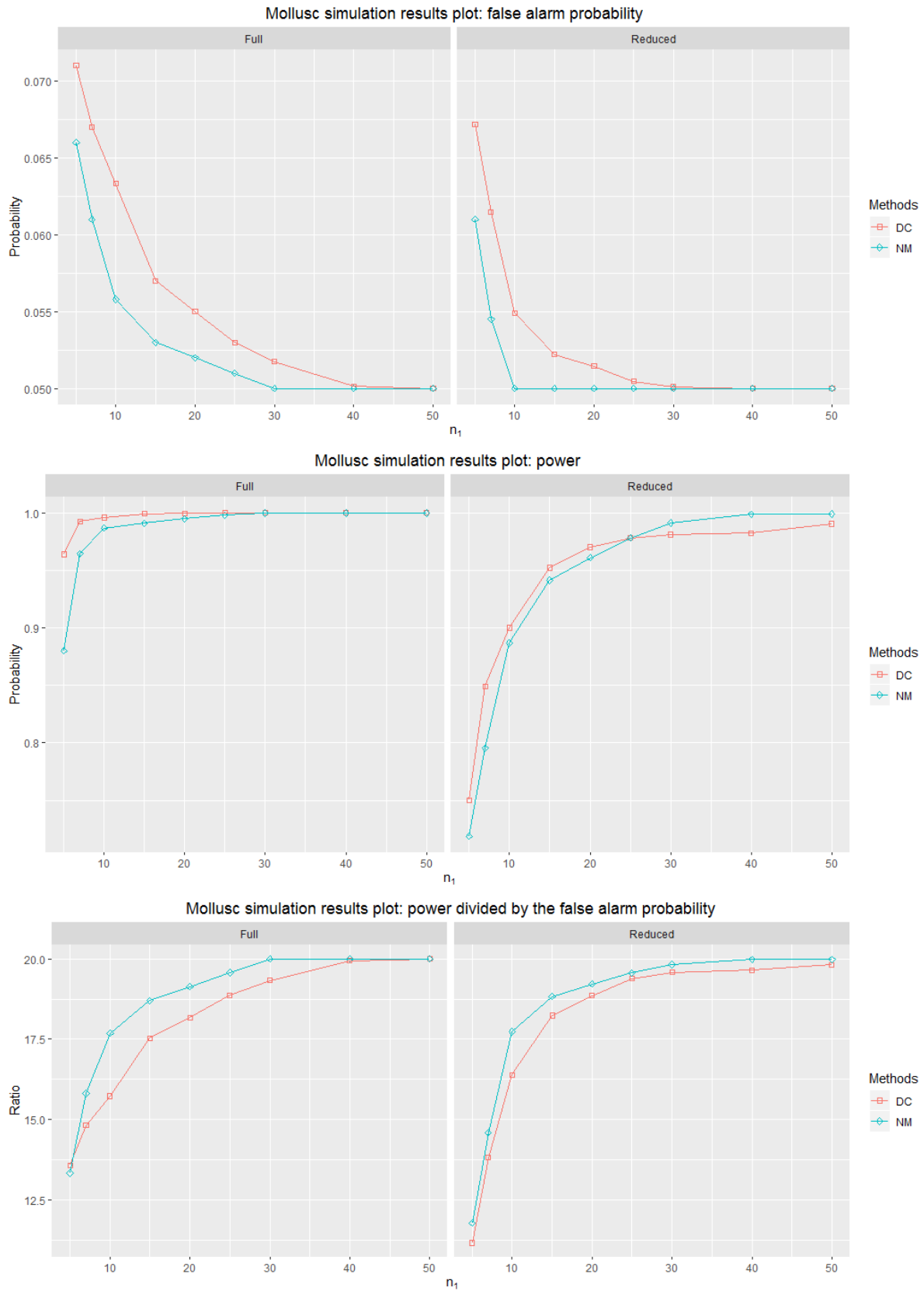


Figure 6.4: Simulation results for the ecological dataset

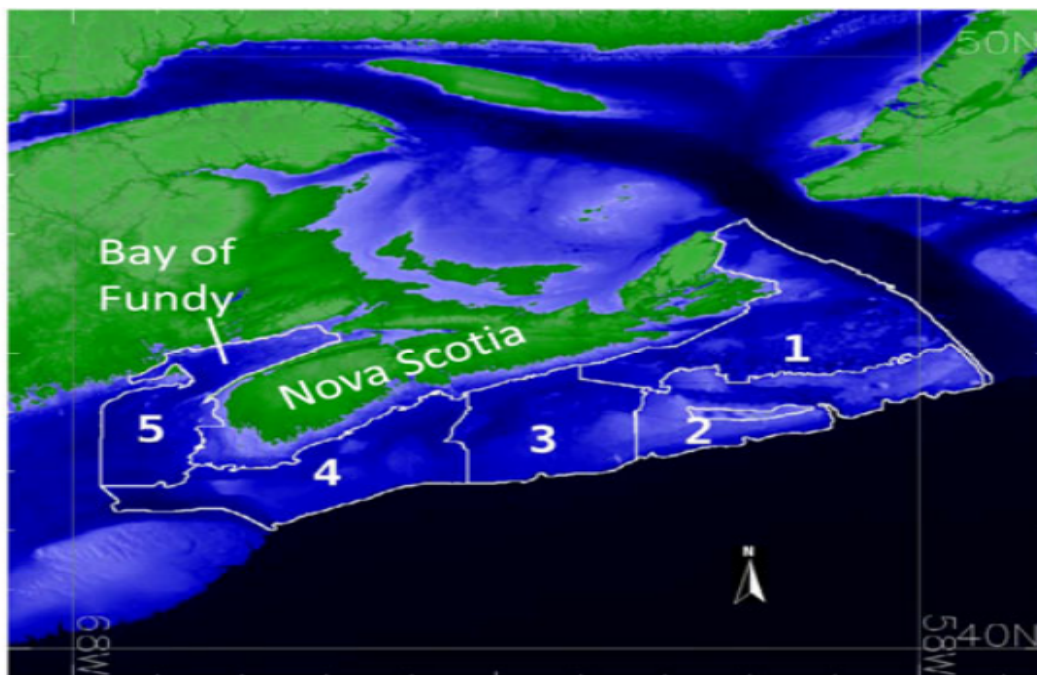


Figure 6.5: Study area and delineation of five sub-areas on the Scotian Shelf off Nova Scotia, Canada

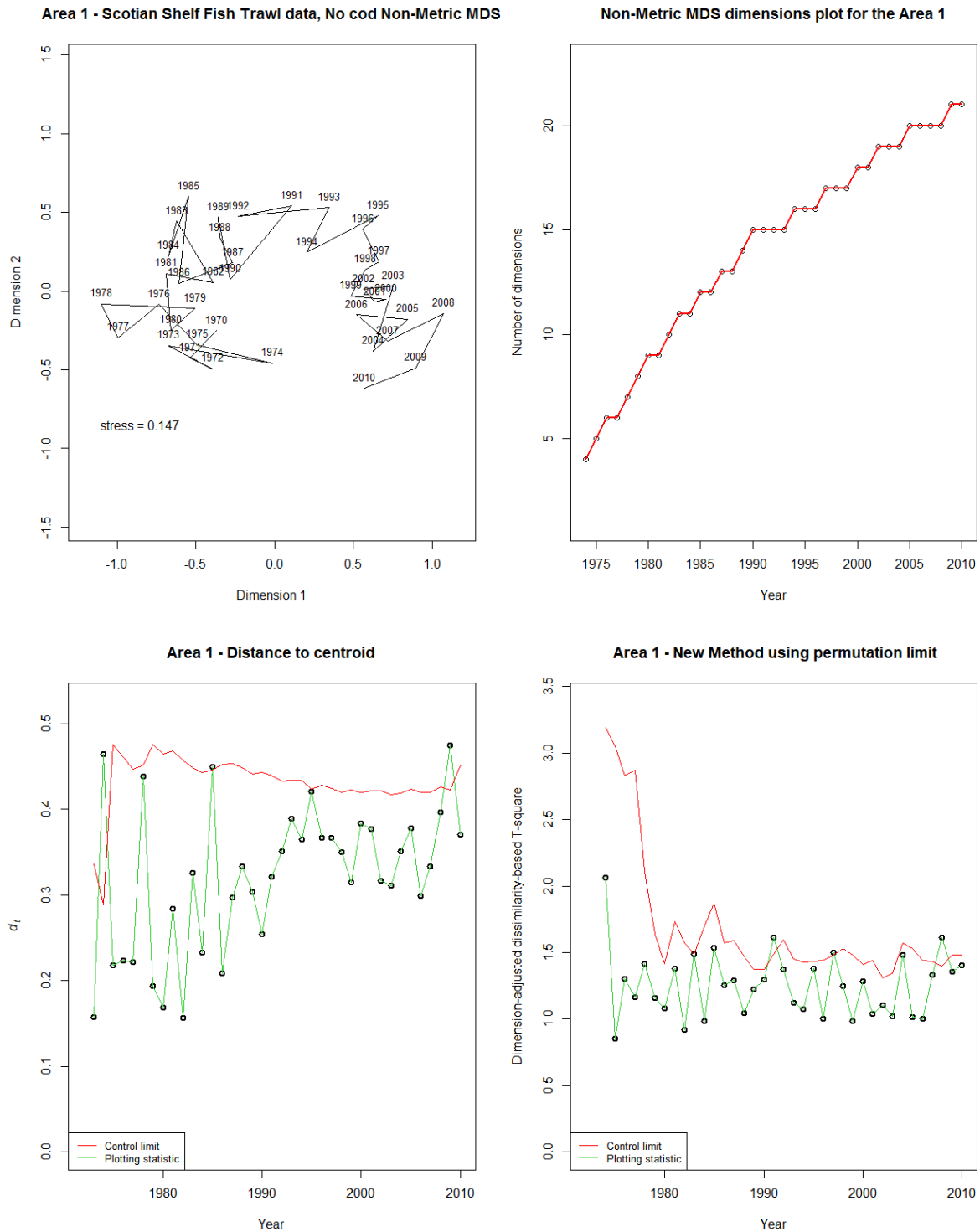


Figure 6.6: Results of the Area of the Scotian Shelf. The figure includes the non-metric MDS 2D plot, the number of dimension retained at any time t^* , the control chart for the area based on the distance to centroid, and the proposed chart together in a figure.

Chapter 7

EWMA control chart for monitoring the mean of a process that is negatively correlated with an auxiliary variable under some ranked sampling schemes

”This is the peer reviewed version of the following article: “Adegoke, N. A., Riaz, M., Sanusi, R. A., Smith, A. N., & Pawley, M. D. (2017). EWMA-type scheme for monitoring location parameter using auxiliary information. *Computers & Industrial Engineering*, 114, 114-129”, which has been published in final form at <https://doi.org/10.1016/j.cie.2017.10.013>. This article may be used for non-commercial purposes in accordance with Elsevier Terms and Conditions for Self-Archiving.”

Control charts are statistical methods used to detect shifts in the location parameter of a process that is monitored over time. Here, we propose an improvement to the performance of the classical exponen-

tially weighted moving average (EWMA) control charts, by making use of auxiliary information that is correlated with the process variable. We present an $EWMA_P$ and its modifications: EWMA-type control charts based on a product estimator where the location parameter of the process is monitored using an auxiliary variable. The charts are developed using different sampling schemes: simple random sampling, ranked set sampling (RSS) and median ranked set sampling (MRSS), and we evaluate their performance using average run length, and other performance measures such as extra quadratic loss and relative average run length. It is observed that the proposed control charts are performing better than the classical EWMA control chart in monitoring shifts in the location parameter of a process, especially when a strong negative correlation exists between the process and the auxiliary variables. They are particularly efficient in detecting small to moderate shifts in the process.

7.1 Introduction

Statistical process control (SPC) involves the application of control charts to monitoring the variation in some quantity of interest, such as a measurement taken from goods produced by an industrial process (May and Spanos, 2006). Variation in the process may be classified into two groups: the natural variation and the assignable variation (Abbasi et al., 2015). The natural variation is an intrinsic component of the production process and is usually unavoidable. The purpose of SPC is to quickly detect the occurrence of the assignable causes of variation in the production process so that the process can be examined and remedial actions are initiated to minimise the number of nonconforming products (Montgomery, 2009). The performance of control charts in identifying the assignable alterations in the production process has been shown to depend on, among other things, the sampling scheme used in their development (Mehmood et al., 2013). Several new techniques and modifications to existing techniques have been proposed. Some of these modifications rely on the utilization of a simple random sampling (SRS) scheme during the

monitoring process; yet, SRS has been shown to be less powerful than some alternative schemes in estimating the population mean. This is due to the fact that estimates from SRS can be highly skewed, causing wide control margins and low control efficiency (Pongpullponsak and Sontisamran, 2013), when compared with some other sampling methods, such as ranked set sampling (RSS) and its modifications (Abujiya and Muttalak, 2004).

McIntyre (1952) originally proposed RSS and demonstrated its superiority over SRS. RSS can improve process control by producing narrower control margins, while being generally cheaper and faster to implement (Pongpullponsak and Sontisamran, 2013). RSS is most useful when the variable of interest is highly difficult or costly to measure, but the ranking can be easily done at negligible cost (Patil et al., 2002; Alamand, Md Sarwar and Sinha, Arun Kumar and Ali, 2016). Even if the ordering is difficult but there is an easily ranked concomitant variable, then it may be used to “judgement order” the original variable (Stokes, 1977; Chen and Shen, 2003). Regardless of how the ranking is done, the efficiency of RSS is expected to be better than SRS, or at least as accurate as SRS with an equal number of quantifications. This is because more information is contained in the RSS than a simple random sampling of the same size (Halls, Lowell K and Dell, 1966; Dell and Clutter, 1972). For further work demonstrating the superiority of RSS (and some modifications) over classical SRS, see Abujiya and Muttalak (2004), Al-Sabah (2010), Mehmood et al. (2013), Pongpullponsak and Sontisamran (2013).

In many cases, direct measurements of the process of interest may be measured alongside some auxiliary variables with which it is correlated. Where available, auxiliary variables may be used to improve the efficiency of control chart schemes, using standard methods such as regression (Riaz, 2008b). Recently, Abbas et al. (2014a) proposed an exponentially weighted moving average (EWMA)-type control chart where the process mean is regressed on the auxiliary variable. This method performed better than both the classical univariate and bivariate EWMA control charts, particularly for detecting small to moderate shifts in the process mean. Thus far, research in this area has focused on cases where the relationship

between the process variable and the auxiliary information is assumed to be positive (Sanusi et al., 2017a). Yet, this is not always the case; the process variable and the auxiliary information may be negatively correlated (Shabbir and Awan, 2016; and others Ahmad, Shabbir and Lin, Zhengyan and Abbasi, Saddam Akber and Riaz, 2012).

Despite the improved efficiency of RSS over SRS, only a few works in the monitoring of the process mean have adopted the use of the RSS, and its modification, especially where the process variable is correlated or observed with another variable (Mehmood et al., 2013; Abujiya, Mu'azu Ramat and Farouk, Abbas Umar and Lee, Muhammad Hisyam and Mohamad, 2013). In this work, we propose an improvement to the performance of the classical EWMA control chart for monitoring shift in the location parameter, using auxiliary information that is negatively correlated with the process variable. Specifically, we present an $EWMA_P$ and its modifications: EWMA-type control charts based on product estimator, where the location parameter of a process is monitored using one auxiliary variable. We develop the charts using different sampling schemes: SRS, RSS and median ranked set sampling (MRSS). We evaluate the proposed charts' performance using average run length (ARL). ARL is the expected number of plotted statistics or samples until we have the first out-of-control signal (Balakrishnan et al., 2010; Haq et al., 2016). We also consider other performance measures such as extra quadratic loss (EQL) and relative average run length (RARL) (Khaliq et al., 2016; Sanusi et al., 2016).

The rest of the chapter is organized as follows: in Section 7.2, we describe the RSS and MRSS procedures, followed by the description of the product estimator, and the efficiency measures of the product estimator from different sampling schemes in Section 7.3. In Section 7.4, we describe the basic structure of the classical EWMA control chart and propose extensions of the classical EWMA control chart to the case where another variable that is negatively correlated with the process variable. Section 7.6 presents the design, performance measure and comparison of the proposed charts with the classical EWMA control chart. We provide illustrative examples in Section 7.7, with a concluding discussion in

7.2 The ranked set sampling (RSS) and median ranked set sampling (MRSS) procedures

Ranked set sampling was introduced by McIntyre (1952), as a more efficient and cost-effective method than the commonly used SRS in the situations where the exact measurement of the units is difficult and expensive, but the visual ordering of the sample units can be done easily (Du and Maceachern, 2008; Alodat et al., 2010). The RSS scheme can be described as follows: Select a random sample of size n^2 units from a given population, and randomly divide the selected n^2 units into n sets, each of size n . The n units within each set are then ordered based on a judgement of relative importance. This can be achieved using a covariate measure that is correlated with the process variable (Patil et al., 2002). Then, the n samples are obtained by choosing the first unit from the first set, the second unit from the second set, and so on. The process may be cycled m ways, until $k = nm$ units are chosen. The nm units constitute the ranked set sample units.

Analytically, let $Y_1, Y_2, Y_3, \dots, Y_n$ be a random sample of size n , from the density function $f(y)$. The simple random sampling estimator of the mean of the study variable Y is given as $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, with standard error $sd(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}}$. Let $Y_{11}, Y_{12}, \dots, Y_{1n}, Y_{21}, Y_{22}, \dots, Y_{2n}, \dots, Y_{n1}, Y_{n2}, \dots, Y_{nn}$ be n independent SRS each of size n . Also, let $Y_{i(1:n)}, Y_{i(2:n)}, \dots, Y_{i(n:n)}$ represent the order statistics of the i th sample (Haq et al., 2013). The measured ranked set sampling units are denoted by $Y_{1(1:n)}, Y_{2(2:n)}, \dots, Y_{n(n:n)}$. Let $g_{(i:n)}(y)$ be the probability density function of the i th order statistic $Y_{(i:n)}$, $i = 1, 2, \dots, n$, from a random sample of

size n . It can be shown that:

$$g_{(i:n)}(y) = n \binom{n-1}{i-1} \{F(y)\}^{i-1} \{1-F(y)\}^{n-i} f(y) \quad -\infty < y < +\infty \quad (7.1)$$

where $F(y)$ is the cumulative distribution of Y (David, H and Nagaraja, 2003; Haq et al., 2013; Alamand, Md Sarwar and Sinha, Arun Kumar and Ali, 2016). The mean and variance of $Y_{(i:n)}$ are given by:

$$\mu_{(i:n)} = \int_{-\infty}^{+\infty} y g_{(i:n)}(y) dy \quad \text{and} \quad \sigma_{(i:n)}^2 = \int_{-\infty}^{+\infty} (y - \mu_{(i:n)})^2 g_{(i:n)}(y) dy \quad (7.2)$$

The ranked set sampling of the mean and variance of the process variable Y are given as (David, H and Nagaraja, 2003; Haq et al., 2013):

$$\bar{Y}_{RSS} = \frac{1}{n} \sum_{i=1}^n Y_{i\{i:n\}}, \quad (7.3)$$

and

$$Var(\bar{Y}_{RSS}) = \frac{1}{n^2} \sum_{i=1}^n \sigma_{i\{i:n\}}^2 = \frac{\sigma_Y^2}{n} - \frac{1}{n^2} \sum_{i=1}^n (\mu_{(i:n)} - \mu_Y)^2 \quad (7.4)$$

$Y_{i\{i:n\}}$ represents the order statistics of the i th sample, and μ_Y is the population mean of Y .

Several modifications of the RSS have been proposed in practice. These include but are not limited to MRSS by Muttalak (1997), while Samawi, Hani M and Ahmed, Mohammad S and Abu-Dayyeh (1996) investigated the extreme ranked set sample (ERSS); also see Samawi and Tawalbeh (2002) for a literature on different double stage RSS. The MRSS is a modification of the RSS which involves: select n^2 samples from a population, and divide this sample as in the case of RSS, to n sets, each of size n . The n units within each set are then ordered based on a judgement of relative importance. The n measurements are obtained depending on whether the set size is odd or even. If the size is even, draw the $(\frac{n}{2})$ th smallest element from the first half, and draw the $(\frac{n+2}{2})$ th smallest element from the remaining sets (second half).

If the size is odd, select the $(\frac{n+1}{2})th$ elements from each ranked set (Abujiya, Mu'azu Ramat and Farouk, Abbas Umar and Lee, Muhammad Hisyam and Mohamad, 2013). The process may be cycled m rounds, until $k = nm$ units are chosen. The $k = nm$ units constitute the median set sample units.

Analytically, the means and variances of Y of the MRSS for both even and odd sample size n are given as (David, H and Nagaraja, 2003; Haq et al., 2013):

$$\bar{Y}_{MRSS}^{Even} = \frac{1}{k} \sum_{j=1}^m \left(\sum_{i=1}^{n/2} Y_{i\{(n/2):n\}j} + \sum_{i=(n/2)+1}^n Y_{i\{(n/2)+1:n\}j} \right) \quad \text{and} \quad \bar{Y}_{MRSS}^{Odd} = \frac{1}{k} \sum_{j=1}^m \sum_{i=1}^n Y_{i\{(n+1)/2:n\}j}$$

\bar{Y}_{MRSS}^{Even} and \bar{Y}_{MRSS}^{Odd} are unbiased estimates of μ_Y , for any symmetric distribution. Their variances, respectively, are given as:

$$Var(\bar{Y}_{MRSS}^{Even}) = E(\bar{Y}_{MRSS}^{Even} - \mu_Y)^2 - \frac{1}{kn} \left(\sum_{i=1}^{n/2} \sigma_{Y\{(n/2):n\}}^2 + \sum_{i=(n/2)+1}^n \sigma_{Y\{(n/2)+1:n\}}^2 \right)$$

$$Var(\bar{Y}_{MRSS}^{Odd}) = E(\bar{Y}_{MRSS}^{Odd} - \mu_Y)^2 - \frac{1}{kn} \sum_{i=1}^n \sigma_{Y\{(n+1)/2:n\}}^2$$

7.3 The product estimator

Assume that Y and X are positive, and can be obtained in paired form for each sample. More so, we assume that Y and X are correlated, and are obtained from bivariate normally distributed random variables, that is, $(Y, X) \sim N_2(\mu_Y, \mu_X, \sigma_Y, \sigma_X, \rho_{XY} : -1 \leq \rho_{XY} \leq 1)$, where N_2 represents the bivariate normal distribution. The product estimator was proposed by Robson (1957), and is given as:

$$\bar{Y}_p = \frac{\bar{Y} \bar{X}}{\mu_X}, \quad (7.5)$$

where \bar{X} and \bar{Y} are the average of the X and Y , respectively, and μ_X is the population mean of X , which is assumed to be known. The mean and MSE of \bar{Y}_p are given as:

$$E(\bar{Y}_p) = \mu_Y \left(1 + \left(\frac{1}{n} \right) \rho_{YX} C_Y C_X \right), \quad (7.6)$$

$$MSE(\bar{Y}_p) = \left(\frac{1}{n} \right) \mu_Y^2 (C_Y^2 + C_X^2 + 2\rho_{YX} C_Y C_X), \quad (7.7)$$

where in Equations (7.6) and (7.7): $C_X = \sigma_X/\mu_X$ and $C_Y = \sigma_Y/\mu_Y$, are the population coefficients of variations for X and Y , respectively, and ρ_{YX} is the correlation coefficient between the variables. Note that sampling with replacement and approximation up to the first order are assumed.

7.3.1 Efficiency of the product estimator based on the different sampling schemes

We study the efficiency of the product estimator based on the different sampling schemes with respect to the conventional estimator of Y . Specifically, we want to find the range of ρ_{XY} , such that:

$$MSE(\bar{Y}^*(\rho_{XY})) \leq Var(\bar{Y}), \quad (7.8)$$

where we have used $MSE(\bar{Y}^*(\rho_{XY}))$ to represent the mean square error of a particular sampling scheme besides the conventional approach, and $Var(\bar{Y})$ is the variance of the conventional approach.

For the conventional product estimator based on simple random sampling, Murthy (1964) showed that when the relationship between Y and X is linear through the origin, and Y is inversely proportional to X , the product estimator will be more efficient than the conventional approach (i.e., \bar{Y}) (Adebola, FB and Adegoke, 2015; Adebola, F.B. and Adegoke, N.A. and Sanusi, 2015). Using the relation in Equation (7.8), Murthy (1964) showed that the \bar{Y}_p in Equation (7.5) for the SRS case is more efficient than the

corresponding variance of \bar{Y} given as σ_Y^2/n , if:

$$\rho_{YX} \leq -\frac{C_X}{2C_Y}. \quad (7.9)$$

An important question in RSS and its modifications where an auxiliary variable has been used is on which of the two variables should the ordering be based (Samawi, Hani M and Muttalak, 1996), as ordering both is not possible in practice, or it might be more straightforward to order one variable rather than the other. We consider the case of correctly applying the schemes on one variable at a particular time, while the ranking on the other variable is done with errors, and obtaining numerically the range of values of ρ_{XY} under which the different sampling schemes considered are more efficient than the conventional approach, for different values of n (i.e., $n = 5$ and 15). In the first case, we draw (y_i, x_i) from bivariate normal distribution, that is, $(Y, X) \sim N_2(\mu_Y = 100, \mu_X = 50, \sigma_Y = 10, \sigma_X = 5, \rho_{XY} : -1 \leq \rho_{XY} \leq 1)$; here, $C_X = C_Y$. For all values of ρ_{XY} , we estimate the standard error from all the different sampling schemes and compare these values against the standard error of the conventional approach. The simulation results in Figure 7.1 show that the condition given in Equation (7.9) for the efficiency of \bar{Y}_p in relation to the conventional estimator becomes $\rho_{XY} \leq -0.5$. As shown on the plot in Figure 7.1, this condition appears to be independent of the sample size n .

However, in all of the other sampling schemes considered in this study, the efficiency of the methods in relation to the conventional approach are functions of the sample size n . Specifically, the bigger the sample size, the bigger the range of the values of ρ_{XY} with which a particular sampling scheme is more efficient than the conventional approach. An example is when we can only rank (i.e., correctly apply RSS) Y while the ranking on X has errors. The exact form of the MSE of the product estimator for the RSS

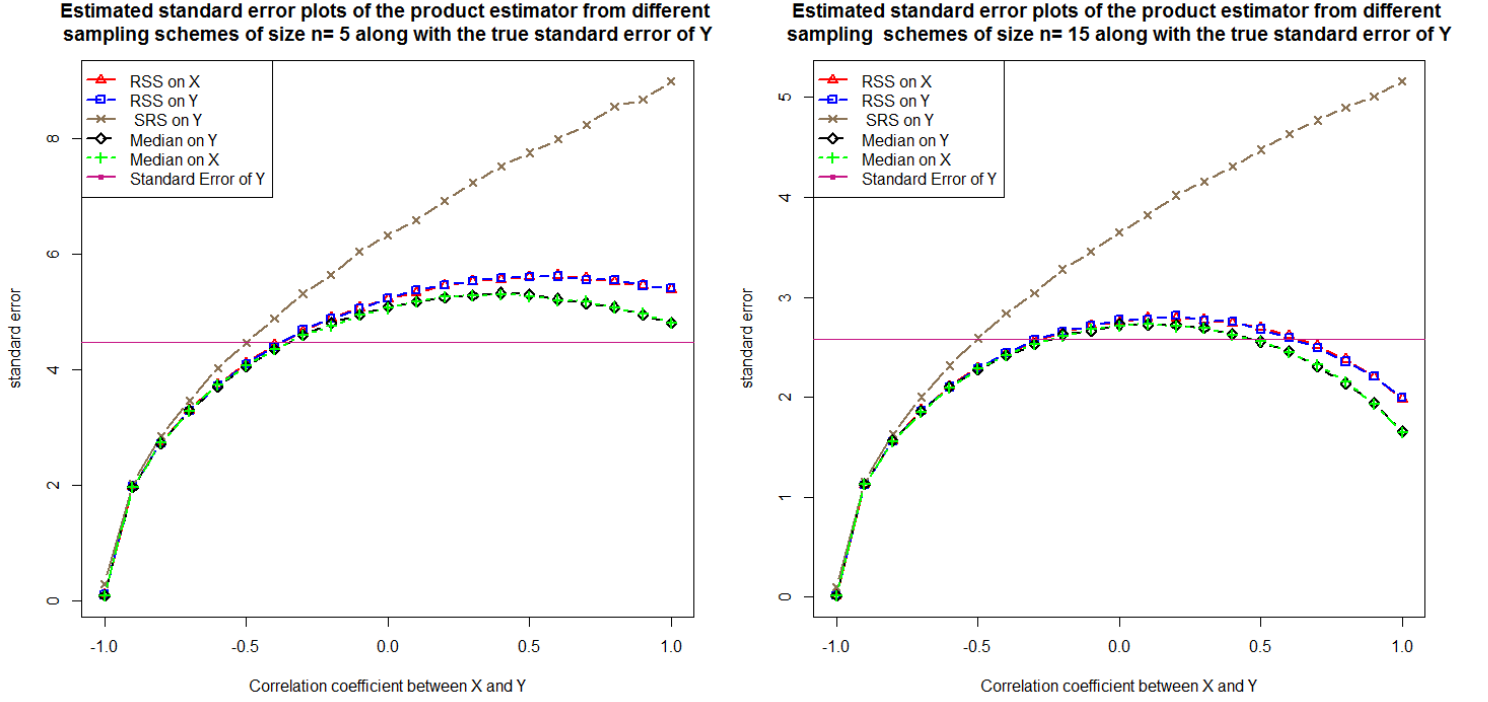


Figure 7.1: The estimated standard error plots of the product estimator from different sampling schemes when $C_X = C_Y$.

on Y , with error sampling on X (i.e, \bar{Y}_{rp1}), is given as:

$$MSE(\bar{Y}_{rp1}) = \left(\frac{1}{n}\right) \mu_Y^2 \left(C_Y^2 + C_X^2 + 2\rho_{YX}C_YC_X - \frac{1}{n} \left[\sum_{i=1}^n Z_{x[i]}^2 + \sum_{i=1}^n Z_{y(i)}^2 + 2 \sum_{i=1}^n Z_{x[i]}Z_{y(i)} \right] \right)$$

where, $Z_{x[i]} = \frac{\mu_{x[i]} - \mu_x}{\mu_x}$ and $Z_{y(i)} = \frac{\mu_{y(i)} - \mu_y}{\mu_y}$. We replace $(.)$ by $[.]$ for the case when ranking is made with errors. That is $MSE(\bar{Y}_{rp1})$ is given as

$$MSE(\bar{Y}_{rp1}) = MSE(\bar{Y}_p) - \left(\frac{\mu_Y^2}{n^2}\right) \left(\sum_{i=1}^n (Z_{x[i]} + Z_{y(i)})^2 \right). \quad (7.10)$$

Using the condition in Equation (7.8), the \bar{Y}_{rp1} will be more efficient than the conventional estimator \bar{Y} if:

$$MSE(\bar{Y}_{rp1}) < Var(\bar{Y}). \quad (7.11)$$

Analytically, this condition holds when:

$$\rho_{YX} \leq k - \frac{C_X}{2C_Y}, \quad (7.12)$$

where k is defined as:

$$k = \frac{1}{2} \left(\frac{\mu_Y^2}{n^2 C_Y C_X} \right) \sum_{i=1}^n \left(Z_{x[i]} + Z_{y(i)} \right)^2. \quad (7.13)$$

Since k in Equation (7.13) is a function of n^2 , then an increase in the sample size n leads to an increase in the range of the values of the correlation with which the sampling scheme is more efficient than the conventional approach. For both RSS and MRSS (when $C_X = C_Y$), on which of the two variables the sampling schemes are applied seems to be less important, because the behavior of their MSE follows the same patterns. Generally, the MRSS, independent of which of the two variables the MRSS is applied, is more efficient than the other sampling schemes.

Lastly, we draw (y_i, x_i) from two different bivariate normal distributions. In the first case, we draw (y_i, x_i) from $(Y, X) \sim N_2(\mu_Y = 100, \mu_X = 50, \sigma_Y = 14, \sigma_X = 5, \rho_{XY} : -1 \leq \rho_{XY} \leq 1)$; here, $C_X < C_Y$. In the second case, we draw (y_i, x_i) from $(Y, X) \sim N_2(\mu_Y = 100, \mu_X = 50, \sigma_Y = 10, \sigma_X = 7, \rho_{XY} : -1 \leq \rho_{XY} \leq 1)$; here, $C_X > C_Y$. When $C_X < C_Y$, for both the MRSS and RSS in Figure 7.2, applying the schemes on the process variable Y leads to more efficiency than applying them on the auxiliary variable. These are shown in the lower MSE for both the RSS on Y and Median on Y in Figure 7.2. Specifically, for all values of ρ_{XY} , there is greater efficiency in the MSE when the RSS and MRSS are applied on Y . However, when $C_X > C_Y$; in Figure 7.3, applying the schemes on the process variable Y leads to less efficiency than applying them on the auxiliary variable. This is also evident in the lower MSE for both the RSS on X and Median on X in Figure 7.3. That is, for all values of ρ_{XY} , there is greater efficiency in the MSE when the RSS and MRSS are applied on X . In both Figure 7.2 and Figure 7.3, an increase in the sample size n also leads to an increase in the range of correlation for which both the RSS and MRSS

are more efficient than the conventional approach. However, increasing the sample size n seems not to have greater effects on the SRS scheme.

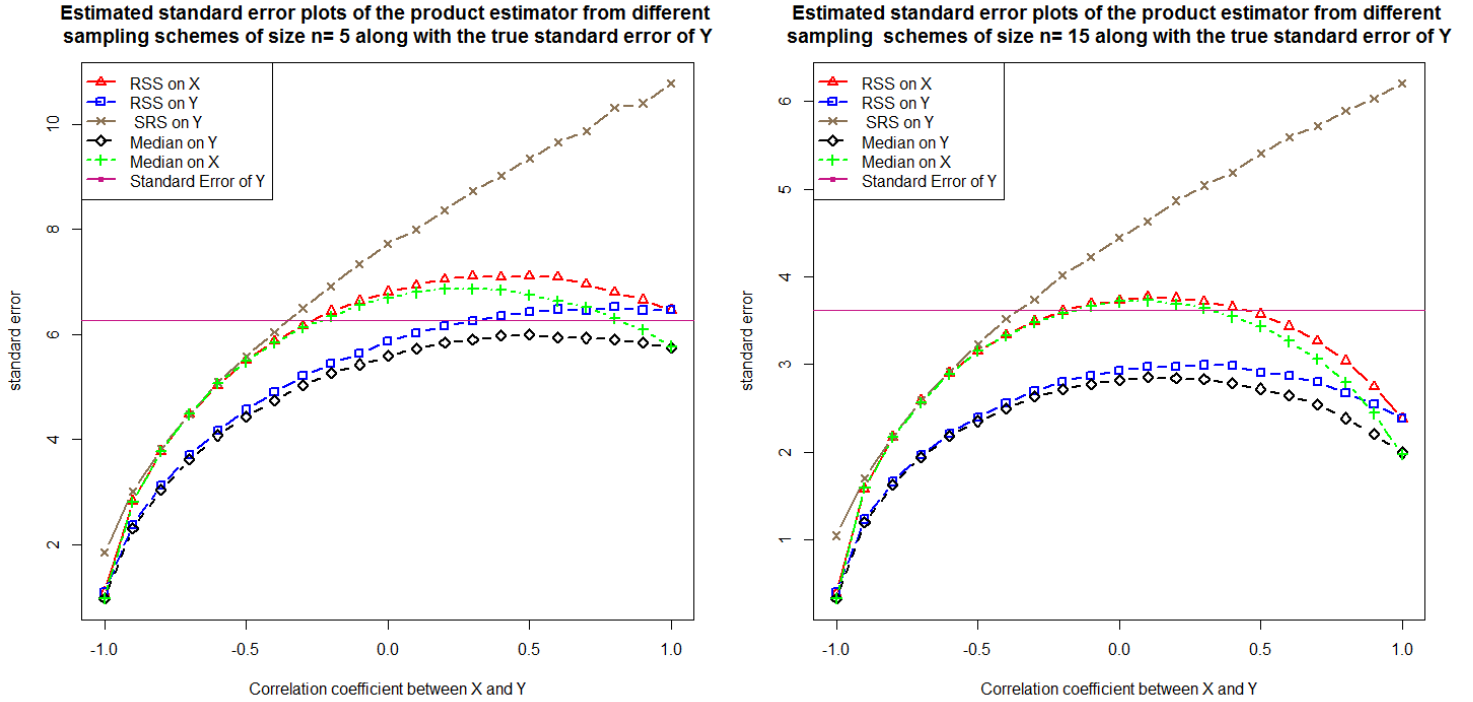


Figure 7.2: The estimated standard error plots of the product estimator from different sampling schemes when $C_X < C_Y$.

In all of our simulation results, irrespective of which of the variables the MRSS or RSS was applied on, the RSS or MRSS is at least as efficient as the product estimator based on the SRS. The numerical results in this section show that the product estimator is much more efficient than the conventional approach when the relationship or correlation between the variables is negative. Specifically, when $-1 \leq \rho_{XY} < -0.5C_X/C_Y$, this result conforms with the conditions in Murthy (1964). However, our simulation results suggest that the range of correlation between X and Y is increased for the product estimator based on the MRSS and RSS.

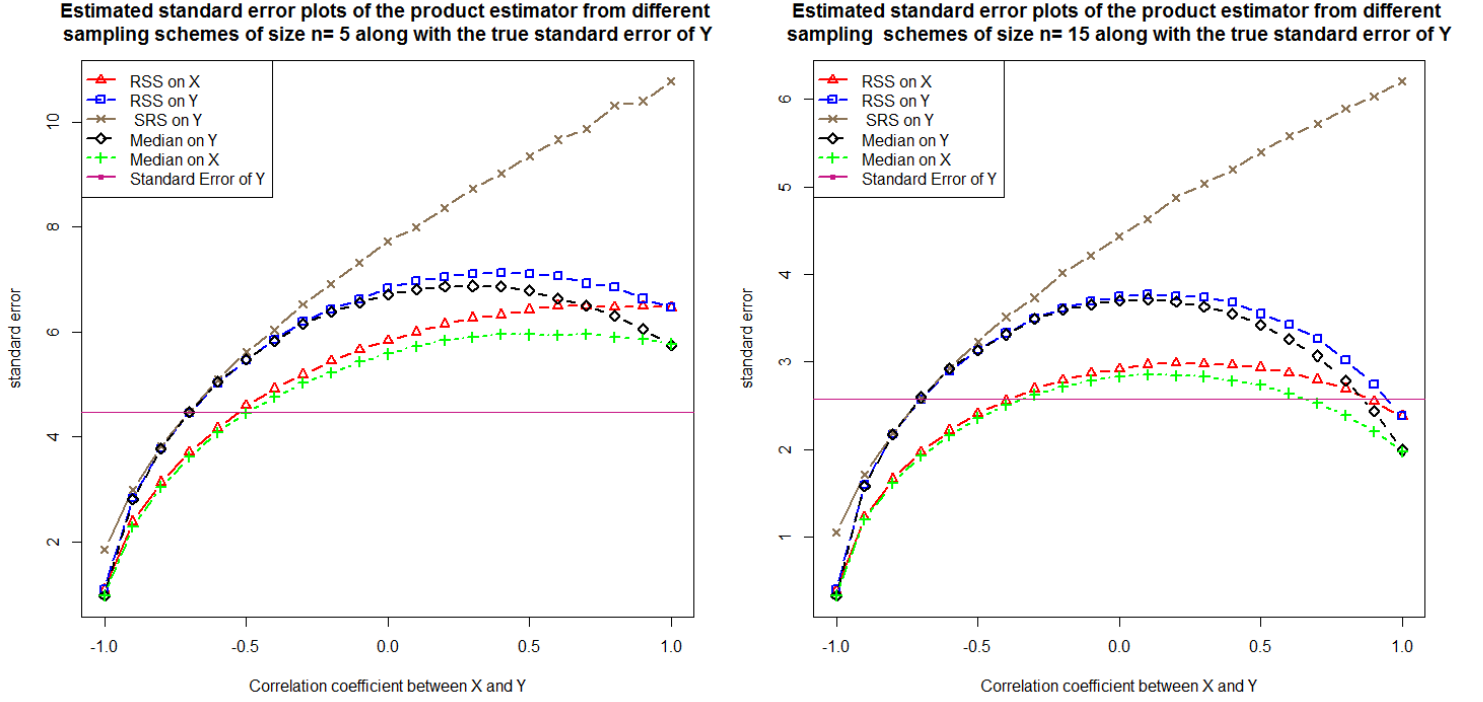


Figure 7.3: The estimated standard error plots of the product estimator from different sampling schemes when $C_X > C_Y$.

7.4 The classical EWMA control chart

Let Y be an independent process variable from a normal distribution with mean μ_Y and variance σ_Y^2 .

The EWMA statistics for monitoring a process shift in the location parameter is given as:

$$W_i = \lambda Y_i + (1 - \lambda)W_{i-1}, \quad (7.14)$$

where λ is the smoothing parameter and it must be chosen such that $0 < \lambda \leq 1$ (Roberts, 1959). W_0 represents the initial value required with $i = 1$ and it is generally chosen to be the process target mean (i.e., $W_0 = \mu_Y$). When μ_Y is unknown, the average of the samples (i.e., $W_0 = \bar{Y}$), is used as the starting value. When choosing the value of λ , it is recommended to use small values to detect small shifts. An EWMA chart with $\lambda = 1.0$ corresponds to a memoryless control chart (Montgomery, 2009). The mean

and variance of W_i are given as:

$$E(W_i) = \mu_Y, \quad (7.15)$$

$$Var(W_i) = \sigma_Y^2 \left\{ \frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i}) \right\}, \quad (7.16)$$

where, μ_Y and σ_Y^2 are the mean and variance of Y , and are estimated when they are unknown.

The control limits of W_i are given as:

$$LCL_i = \mu_Y - L\sigma_Y \sqrt{\left\{ \frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i}) \right\}}, \quad (7.17)$$

$$CL = \mu_Y,$$

$$UCL_i = \mu_Y + L\sigma_Y \sqrt{\left\{ \frac{\lambda}{2-\lambda} (1 - (1-\lambda)^{2i}) \right\}}. \quad (7.18)$$

The EWMA chart is very effective in detecting a small process shift. The design parameters of EWMA chart are L and the value of λ (Roberts, 1959; Cheng, Smiley W. and Thaga, 2006). These parameters are chosen to give the desired in-control ARL performance for the chart (Montgomery, 2009; Teoh, Wei Lin and Khoo, Michael BC and Castagliola, Philippe and Lee, 2016).

The term $(1 - (1-\lambda)^{2i})$ in Equation (7.17) and Equation (7.18), approaches unity as i gets larger (May and Spanos, 2006), and the LCL and UCL approach the steady state values given by (Sanusi et al., 2017b):

$$LCL = \mu_Y - L\sigma_Y \sqrt{\left(\frac{\lambda}{2-\lambda} \right)}, \quad (7.19)$$

$$UCL = \mu_Y + L\sigma_Y \sqrt{\left(\frac{\lambda}{2-\lambda} \right)}. \quad (7.20)$$

The control limits given in Equation (7.19) and Equation (7.20), are called constant limits. The limits

given in Equation (7.17) and Equation (7.18), are called the time varying limits. The time varying limits are more robust than the constant limits in detecting the initial out-of-control conditions (Abbas et al., 2014a).

7.5 The proposed charts

7.5.1 The proposed $EWMA_P$ control chart based on SRS

The proposed modified estimator is given as:

$$W_i = \lambda \bar{Y}_{pi} + (1 - \lambda)W_{i-1}, \quad (7.21)$$

where, \bar{Y}_{pi} is the value of the statistic \bar{Y}_p (given in Equation (7.5)) for the i_{th} sample observation, and W_{i-1} represents the past information, with an initial value given as: $W_0 = E(\bar{Y}_p)$. The varying control limits for the proposed control chart: $EWMA_P$, are given as:

$$LCL_i = E(\bar{Y}_p) - L_c \sqrt{MSE(\bar{Y}_p)} \sqrt{\left\{ \frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right\}}, \quad (7.22)$$

$$CL = E(\bar{Y}_p),$$

$$UCL_i = E(\bar{Y}_p) + L_c \sqrt{MSE(\bar{Y}_p)} \sqrt{\left\{ \frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right\}}. \quad (7.23)$$

As the product estimator is a biased estimator of the mean of Y , the design parameter L_c is adjusted to accommodate for the biasedness of the estimator, and it determines the width of the control limits. The value of L_c is chosen along with the smoothing parameter λ to give the desired in-control ARL of the chart.

7.5.2 The proposed EWMA_{P_{RY}} and EWMA_{P_{RX}} control charts based on RSS

We consider the case when we can correctly apply the ranked set sampling scheme on one variable at a particular time, while the ranking on the other variable is applied erroneously, and propose the EWMA_{P_{RY}} and EWMA_{P_{RX}} control charts based on RSS. We use EWMA_{P_{RY}} for when we can correctly rank on the process variable Y with error ranking on the auxiliary variable X , and EWMA_{P_{RX}} for when we can correctly rank X with error ranking on the process variable Y .

Assuming that we can only correctly rank Y , while the ordering on X is applied erroneously, $(Y_{(r)}, X_{[r]})$; where $r = i : n$, is an i_{th} order elements in the n_{th} sample for the process variable Y , and the i_{th} judgemental ranking in the n_{th} sample for X . The product estimator for the RSS data with errors in ordering of X for the m_{th} circle (without loss of generality, we use $m = 1$) is given by:

$$\bar{Y}_{rp1} = \frac{\bar{Y}_{(n)m} \bar{X}_{[n]m}}{\mu_X}, \quad (7.24)$$

where $\bar{Y}_{(n)m} = \frac{1}{n} \sum_{i=1}^n Y_{(r)m}$ and $\bar{X}_{[n]m} = \frac{1}{n} \sum_{i=1}^n X_{[r]m}$ are the means of the perfect RSS for the variable Y , and error in ranking X , respectively. The mean and MSE of \bar{Y}_{rp1} in Equation (7.24) are given as:

$$E(\bar{Y}_{rp1}) = \mu_Y \left(1 + \frac{1}{n} \rho_{Y(r)X[r]} C_{Y(r)} C_{X[r]} \right), \quad (7.25)$$

$$MSE(\bar{Y}_{rp1}) = \left(\frac{1}{n} \right) \mu_Y^2 (C_{Y(r)}^2 + C_{X[r]}^2 + 2\rho_{Y(r)X[r]} C_{Y(r)} C_{X[r]}). \quad (7.26)$$

where $C_{X[r]} = \frac{\sigma_{X[r]}}{\mu_X}$ and $C_{Y(r)} = \frac{\sigma_{Y(r)}}{\mu_Y}$ are the coefficients of variations for $X_{[r]}$ and $Y_{(r)}$, respectively, $\sigma_{X[r]} = \sqrt{\sigma_{X[r]}^2}$ and $\sigma_{Y(r)} = \sqrt{\sigma_{Y(r)}^2}$. In addition, $\sigma_{X[r]}^2 = E[(X_{[r]} - E(X_{[r]}))^2]$ and $\sigma_{Y(r)}^2 = E[(Y_{(r)} - E(Y_{(r)}))^2]$ are the variances of $X_{[r]}$ and $Y_{(r)}$, respectively.

The modified EWMA statistic in this case is given as:

$$W_i^1 = \lambda \bar{Y}_{rp1i} + (1 - \lambda) W_{i-1}^1, \quad (7.27)$$

where, \bar{Y}_{rp1i} is the value of the statistic \bar{Y}_{rp1} for the i_{th} sample observation, and W_{i-1}^1 represents the past information, with the initial value given as: $W_0^1 = E(\bar{Y}_{rp1})$. The time varying control limits of the proposed control chart $EWMA_{PRY}$ are given as:

$$LCL_i = E(\bar{Y}_{rp1}) - L_c \sqrt{MSE(\bar{Y}_{rp1})} \sqrt{\left(\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right)}, \quad (7.28)$$

$$CL = E(\bar{Y}_{rp1}),$$

$$UCL_i = E(\bar{Y}_{rp1}) + L_c \sqrt{MSE(\bar{Y}_{rp1})} \sqrt{\left(\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right)}. \quad (7.29)$$

However, when we can only correctly order X , while the ordering on Y is applied with errors, then $(Y_{[r]}, X_{(r)})$ is the i_{th} judgemental ordering in the n_{th} sample for Y , and i_{th} order elements in the n_{th} sample for X . The product estimator for the RSS data with errors in ordering of Y for the m_{th} circle is given by:

$$\bar{Y}_{rp2} = \frac{\bar{Y}_{[n]m} \bar{X}_{(n)m}}{\mu_X}, \quad (7.30)$$

where $\bar{Y}_{[n]m} = \frac{1}{n} \sum_{i=1}^n Y_{[r]m}$ and $\bar{X}_{(n)m} = \frac{1}{n} \sum_{i=1}^n X_{(r)m}$ are the means of the perfect RSS for the variable X , and error in ranking Y , respectively. The mean and MSE of \bar{Y}_{rp2} in Equation (7.30) are given as:

$$E(\bar{Y}_{rp2}) = \mu_Y \left(1 + \frac{1}{n} \rho_{Y_{[r]}X_{(r)}} C_{Y_{[r]}} C_{X_{(r)}} \right), \quad (7.31)$$

$$MSE(\bar{Y}_{rp2}) = \left(\frac{1}{n} \right) \mu_Y^2 (C_{Y_{[r]}}^2 + C_{X_{(r)}}^2 + 2 \rho_{Y_{[r]}X_{(r)}} C_{Y_{[r]}} C_{X_{(r)}}). \quad (7.32)$$

where $C_{X_{(r)}} = \frac{\sigma_{X_{(r)}}}{\mu_X}$ and $C_{Y_{[r]}} = \frac{\sigma_{Y_{[r]}}}{\mu_Y}$ are the coefficients of variations for $X_{(r)}$ and $Y_{[r]}$, respectively, $\sigma_{X_{(r)}} = \sqrt{\sigma_{X_{(r)}}^2}$ and $\sigma_{Y_{[r]}} = \sqrt{\sigma_{Y_{[r]}}^2}$. In addition, $\sigma_{X_{(r)}}^2 = E[(X_{(r)} - E(X_{(r)}))^2]$ and $\sigma_{Y_{[r]}}^2 = E[(Y_{[r]} - E(Y_{[r]}))^2]$ are the variances of $X_{(r)}$ and $Y_{[r]}$, respectively.

The modified EWMA statistic in this case is given as:

$$W_i^2 = \lambda \bar{Y}_{rp2i} + (1 - \lambda) W_{i-1}^2, \quad (7.33)$$

where \bar{Y}_{rp2i} is the value of the statistic \bar{Y}_{rp2} for the i_{th} sample observation, and W_{i-1}^2 represents the past information, with the initial value given as: $W_0^2 = E(\bar{Y}_{rp2})$. The time varying control limits of the proposed control chart $EWMA_{PRX}$ are given as:

$$LCL_i = E(\bar{Y}_{rp2}) - L_c \sqrt{MSE(\bar{Y}_{rp2})} \sqrt{\left(\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right)}, \quad (7.34)$$

$$CL = E(\bar{Y}_{rp2}),$$

$$UCL_i = E(\bar{Y}_{rp2}) + L_c \sqrt{MSE(\bar{Y}_{rp2})} \sqrt{\left(\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right)}. \quad (7.35)$$

7.5.3 The proposed $EWMA_{PMY}$ and $EWMA_{PMX}$ control charts based on MRSS

We also consider when we can correctly apply the MRSS on one variable at a particular time, while the ranking on the other variable is applied erroneously, and propose the $EWMA_{PMY}$ and $EWMA_{PMX}$ control charts based on the MRSS scheme. We use $EWMA_{PMY}$ for when we can correctly apply MRSS on Y with error ranking on X , and $EWMA_{PMX}$ for when we can correctly apply MRSS on X with error ranking on Y .

The product estimator of the MRSS sample with error in ranking of X for the mth circle is given by:

$$\bar{Y}_{rp1} = \frac{\bar{Y}_{(n)m} \bar{X}_{[n]m}}{\mu_X}, \quad (7.36)$$

where $\bar{Y}_{(n)m} = \frac{1}{n} \sum_{i=1}^n Y_{(r_{md})m}$ and $\bar{X}_{[n]m} = \frac{1}{n} \sum_{i=1}^n X_{[r_{md}]m}$ are the means of the correct MRSS for the variable Y , and error in ranking X in the m_{th} circle, respectively. Using MRSS $(Y_{(r_{md})}, X_{[r_{md}]})$; where $r_{md} = i : n$, is an i_{th} MRSS observation in the n_{th} sample for the variable Y and the i_{th} judgemental ordering in the n_{th} sample for the auxiliary variable X . The mean and MSE of the estimator are given as:

$$E(\bar{Y}_{rp3}) = \mu_Y \left(1 + \frac{1}{n} \rho_{Y(r_{md})X[r_{md}]} C_{Y(r_{md})} C_{X[r_{md}]}\right), \quad (7.37)$$

$$MSE(\bar{Y}_{rp3}) = \left(\frac{1}{n}\right) \mu_Y^2 (C_{Y(r_{md})}^2 + C_{X[r_{md}]}^2 + 2\rho_{Y(r_{md})X[r_{md}]} C_{Y(r_{md})} C_{X[r_{md}]}). \quad (7.38)$$

where $C_{X[r_{md}]} = \frac{\sigma_{X[r_{md}]}}{\mu_X}$ and $C_{Y(r_{md})} = \frac{\sigma_{Y(r_{md})}}{\mu_Y}$ are the coefficients of variations for $X_{[r_{md}]}$ and $Y_{(r_{md})}$, respectively, $\sigma_{X[r_{md}]} = \sqrt{\sigma_{X[r_{md}]}^2}$ and $\sigma_{Y(r_{md})} = \sqrt{\sigma_{Y(r_{md})}^2}$. In addition, $\sigma_{X[r_{md}]}^2 = E(X_{[r_{md}]} - E(X_{[r_{md}]})^2$ and $\sigma_{Y(r_{md})}^2 = E(Y_{(r_{md})} - E(Y_{(r_{md})}))^2$ are the variances of $X_{[r_{md}]}$ and $Y_{(r_{md})}$, respectively.

The modified EWMA statistic in this case is given as:

$$W_i^3 = \lambda \bar{Y}_{rp3i} + (1 - \lambda) W_{i-1}^3, \quad (7.39)$$

where \bar{Y}_{rp3i} is the value of the statistic \bar{Y}_{rp3} for the i_{th} sample observation, and W_{i-1}^3 represents the past information, with the initial value given as: $W_0^3 = E(\bar{Y}_{rp3})$. The time varying control limits of the

proposed control chart EWMA_{PMY} are given as:

$$LCL_i = E(\bar{Y}_{rp3}) - L_c \sqrt{MSE(\bar{Y}_{rp3})} \sqrt{\left(\frac{\lambda}{2-\lambda}(1 - (1-\lambda)^{2i})\right)}, \quad (7.40)$$

$$CL = E(\bar{Y}_{rp3}),$$

$$UCL_i = E(\bar{Y}_{rp3}) + L_c \sqrt{MSE(\bar{Y}_{rp3})} \sqrt{\left(\frac{\lambda}{2-\lambda}(1 - (1-\lambda)^{2i})\right)}. \quad (7.41)$$

When we can only correctly apply MRSS on X , while the ordering on Y is with errors, then $(Y_{[r_{md}]}, X_{(r_{md})})$ is the i_{th} judgemental ordering in the n_{th} sample for Y , and i_{th} MRSS elements in the n_{th} sample for X . The product estimator of the MRSS sample with error in the ranking of Y for the m_{th} circle is given by:

$$\bar{Y}_{rp1} = \frac{\bar{Y}_{[n]m} \bar{X}_{(n)m}}{\mu_X}, \quad (7.42)$$

where $\bar{Y}_{[n]m} = \frac{1}{n} \sum_{i=1}^n Y_{[r_{md}]m}$ and $\bar{X}_{(n)m} = \frac{1}{n} \sum_{i=1}^n X_{(r_{md})m}$ are the means of the correct MRSS for the variable X , and error in ranking Y in the m_{th} circle, respectively. Using MRSS $(Y_{[r_{md}]}, X_{(r_{md})})$; where $r_{md} = i : n$, is an i_{th} MRSS observation in the n_{th} sample for the variable X and the i_{th} judgemental ordering in the n_{th} sample for the auxiliary variable Y . The mean and MSE of the estimator are given as:

$$E(\bar{Y}_{rp4}) = \mu_Y \left(1 + \frac{1}{n} \rho_{Y_{[r_{md}]} X_{(r_{md})}} C_{Y_{[r_{md}]}} C_{X_{(r_{md})}}\right), \quad (7.43)$$

$$MSE(\bar{Y}_{rp4}) = \left(\frac{1}{n}\right) \mu_Y^2 (C_{Y_{[r_{md}]}}^2 + C_{X_{(r_{md})}}^2 + 2\rho_{Y_{[r_{md}]} X_{(r_{md})}} C_{Y_{[r_{md}]}} C_{X_{(r_{md})}}). \quad (7.44)$$

where $C_{X_{(r_{md})}} = \frac{\sigma_{X_{(r_{md})}}}{\mu_X}$ and $C_{Y_{[r_{md}]}} = \frac{\sigma_{Y_{[r_{md}]}}}{\mu_Y}$ are the coefficients of variations for $X_{(r_{md})}$ and $Y_{[r_{md}]}$, respectively, $\sigma_{X_{(r_{md})}} = \sqrt{\sigma_{X_{(r_{md})}}^2}$ and $\sigma_{Y_{[r_{md}]}} = \sqrt{\sigma_{Y_{[r_{md}]}}^2}$. In addition, $\sigma_{X_{(r_{md})}}^2 = E(X_{(r_{md})} - E(X_{(r_{md})}))^2$ and $\sigma_{Y_{[r_{md}]}}^2 = E(Y_{[r_{md}]} - E(Y_{[r_{md}]}))^2$ are the standard deviations of $X_{(r_{md})}$ and $Y_{[r_{md}]}$, respectively.

The modified EWMA statistic in this case is given as:

$$W_i^4 = \lambda \bar{Y}_{rp4i} + (1 - \lambda) W_{i-1}^4, \quad (7.45)$$

where \bar{Y}_{rp4i} is the value of the statistic \bar{Y}_{rp4} for the i_{th} sample observation, and W_{i-1}^4 represents the past information, with the initial value given as: $W_0^4 = E(\bar{Y}_{rp4})$. The time varying control limits of the the proposed control chart $EWMA_{PMX}$ are given as:

$$LCL_i = E(\bar{Y}_{rp4}) - L_c \sqrt{MSE(\bar{Y}_{rp4})} \sqrt{\left(\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right)}, \quad (7.46)$$

$$CL = E(\bar{Y}_{rp4}),$$

$$UCL_i = E(\bar{Y}_{rp4}) + L_c \sqrt{MSE(\bar{Y}_{rp4})} \sqrt{\left(\frac{\lambda}{2 - \lambda} (1 - (1 - \lambda)^{2i}) \right)}. \quad (7.47)$$

where L_c determines the width of control limits, and it is chosen along with the smoothing parameter λ to give a desired in-control ARL of the chart.

7.6 Performance measures and comparison

We perform a complete evaluation of the proposed charts in terms of average run length (ARL). When the process is in control, a control chart is expected to have a large ARL so that the chart is rarely expected to signal. Conversely, when the process is out of control, a chart is supposed to have a small ARL as it should identify the out-of-control condition promptly (Montgomery, 2009). Specifically, the performance measures are made by creating the charts to have the same in-control ARL and then comparing the out-of-control ARLs for a given process shift. For the given process shift, the chart with the smaller out-of-control ARL is regularly considered to perform better (Jones et al., 2001).

We assume that the parameters are unknown, and the proposed charts are employed in a two-phase

procedure (Jensen et al., 2006); in Phase I, they are used to study a historical reference sample. The modified EWMA means and mean square errors are estimated from the Phase I data, and control limits are obtained for use in Phase II. During Phase II, future data points are checked for deviations from the in-control state. While successive observed future values (or statistics) are within the in-control Phase I limits, the process is believed to be in control. Values that are outside the control limits are indications that there may be an assignable cause, and remedial responses are required in the process. We choose the proposed charts' design parameters L_c and λ to fix the in-control ARL (ARL_0) to 500; this makes it reasonable to compare the proposed charts with other existing charts of the same ARL_0 . The values of L_c when the correlation between the variables (i.e., ρ_{XY}) are $-0.25, -0.5, -0.75, -0.95$, and $\lambda = 0.05$, are 2.641, 2.643, 2.646 and 2.650, respectively.

The following procedures are used in our simulations

- Phase I

- (i) For each sampling scheme, we generate n pairs of bivariate normally distributed random variables (y_i, x_i) from $(Y, X) \sim N_2(\mu_Y, \mu_X, \sigma_Y, \sigma_X, \rho_{XY})$. In line with the results in Section 7.3.1, the following conditions are used: we restrict the range of the correlation between the variables to $-0.95 \leq \rho_{XY} \leq -0.25$, and used three different choices of n ; that is, $n = 5, 10$ and 15 . We draw the data sets from three different choices of the bivariate normal distributions. In case I, we obtained (y_i, x_i) from $(Y, X) \sim N_2(\mu_Y = 100, \mu_X = 50, \sigma_Y = 10, \sigma_X = 5, \rho_{XY})$; here $C_X = C_Y$. In the second case, we obtained (y_i, x_i) from $(Y, X) \sim N_2(\mu_Y = 100, \mu_X = 50, \sigma_Y = 14, \sigma_X = 5, \rho_{XY})$; here $C_X < C_Y$. In the third case, we obtained (y_i, x_i) from $(Y, X) \sim N_2(\mu_Y = 100, \mu_X = 50, \sigma_Y = 10, \sigma_X = 7, \rho_{XY})$; here $C_X > C_Y$.
- (ii) We compute the product estimators from the n bivariate samples.
- (iii) We repeat steps *i* and *ii* 5000 times, and compute the estimated parameters, that is, the

estimated mean and standard error of the product estimators. These are used to set the control limits in phase II.

- Phase II

- (i) At each time i , we generate a pair of random variables (y_i, x_i) from bivariate population with $(Y, X) \sim N_2(\mu_1 = \mu_Y + \delta * (\sigma_Y / \sqrt{n}), \mu_X, \sigma_Y, \sigma_X, \rho_{XY} : -0.95 \leq \rho_{XY} \leq -0.25)$ for each sampling scheme, where δ is the size of the shift that we were interested in detecting.
- (ii) We compute the product estimators from the n Phase II bivariate samples, call this Y_{pi} .
- (iii) Using the estimated parameters from phase I, we compute the modified EWMA control charts statistics, and compare the modified EWMA statistics against the estimated control limits.
- (iv) We repeat steps (i-iii) and record the iteration number (or RL) that gives the first out-of-control signal.

- We repeat Phase I and Phase II processes K times; where K is the number of the simulation trials.

We used $K = 100,000$ in our case.

- The average of the run lengths across simulations (i.e., ARL) was reported.

The results in Tables 7.1 to 7.3 (for $n = 5$), show that the out-of-control ARLs (i.e., when $\delta > 0$) of the charts depend on the size of ρ_{XY} . The larger the absolute size of the correlation between the study variable and the auxiliary variable, the better the efficiency of the modified EWMA control charts over the classical EWMA control chart, especially for smaller values of the shifts (i.e., δ). Also, the out-of-control ARLs of the charts depend on the size of the coefficients of variation of X relative to Y . For example, when $C_X = C_Y$, the efficiency of the proposed charts over the classical chart are pronounced when $\rho_{XY} \leq -0.5$. In all of the cases considered, the modified control charts based on the MRSS, followed

by the RSS schemes, are more efficient in detecting the out-of-control behavior than the modified control charts based on the simple random sampling and the classical control chart.

Table 7.1: The ARL values of the modified EWMA control charts along with the ARL values of classical EWMA control chart when $C_X = C_Y$, and $n = 5$ for $\rho_{XY} = -0.25, -0.5, -0.75$ and -0.95 .

Method	ρ_{XY}	δ										
		0	0.25	0.5	0.75	1	1.5	2	2.5	3	4	5
Classical EWMA	-0.25	500.01	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		499.98	106.41	34.27	16.71	10.3	5.26	3.33	2.38	1.86	1.32	1.1
EWMA _{PRY}		500.39	85.46	26.35	13.28	8.16	4.21	2.7	1.97	1.57	1.18	1.04
EWMA _{PRX}		500.01	88.85	26.86	13.4	8.24	4.22	2.71	1.96	1.55	1.15	1.03
EWMA _{PMY}		500.43	85.56	26.04	12.92	7.99	4.11	2.65	1.94	1.53	1.16	1.04
EWMA _{PMX}		500.12	83.41	25.71	12.88	7.93	4.08	2.62	1.9	1.51	1.13	1.02
Classical EWMA	-0.50	500	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		499.94	77.25	23.66	11.84	7.35	3.81	2.45	1.81	1.45	1.11	1.02
EWMA _{PRY}		500	68.32	20.62	10.29	6.41	3.34	2.18	1.62	1.33	1.07	1.01
EWMA _{PRX}		500.01	65.64	20.04	10.17	6.3	3.28	2.14	1.6	1.29	1.05	1
EWMA _{PMY}		499.96	64.13	19.78	9.96	6.17	3.25	2.13	1.6	1.3	1.06	1.01
EWMA _{PMX}		500.01	65.08	19.86	10.02	6.2	3.23	2.11	1.57	1.28	1.04	1
Classical EWMA	-0.75	499.02	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		500	43.34	13.16	6.70	4.18	2.25	1.54	1.22	1.07	1	1
EWMA _{PRY}		500.01	39.81	12.19	6.23	3.91	2.13	1.47	1.18	1.06	1	1
EWMA _{PRX}		500.96	40.68	12.34	6.23	3.9	2.11	1.45	1.16	1.04	1	1
EWMA _{PMY}		499.69	39.17	12.03	6.15	3.86	2.11	1.46	1.18	1.05	1	1
EWMA _{PMX}		500	39.99	12.15	6.17	3.85	2.09	1.44	1.15	1.04	1	1
Classical EWMA	-0.95	500	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		500.01	11.17	3.61	1.98	1.38	1.03	1	1	1	1	1
EWMA _{PRY}		499.97	10.78	3.47	1.91	1.34	1.02	1	1	1	1	1
EWMA _{PRX}		500	10.86	3.46	1.91	1.34	1.02	1	1	1	1	1
EWMA _{PMY}		500.43	10.77	3.46	1.91	1.34	1.02	1	1	1	1	1
EWMA _{PMX}		500	10.86	3.48	1.9	1.33	1.02	1	1	1	1	1

Furthermore, the choice of which variable the MRSS or RSS is to be applied also depends on the size of the estimated value of C_X relative to C_Y . When $C_X = C_Y$ in Table 7.1, it is less important whether the schemes are applied to X or Y . However, in Table 7.2, where $C_X < C_Y$, applying the scheme on the process variable Y gives more efficiency to out-of-control performance than applying the schemes on the auxiliary variable X for both RSS and MRSS. Specifically, our simulation results show that applying the RSS or MRSS on the variable with the larger value of the coefficient of variations gives better efficiency to the out-of-control performance. These results are supported by the output in Table 7.3, where $C_X > C_Y$

Table 7.2: The ARL values of the modified EWMA control charts along with the ARL values of classical EWMA control chart when $C_X < C_Y$, and $n = 5$ for $\rho_{XY} = -0.25, -0.5, -0.75$ and -0.95 .

Method	ρ_{XY}	δ										
		0	0.25	0.5	0.75	1	1.5	2	2.5	3	4	5
Classical EWMA	-0.25	500.34	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		500.01	85.5	26.44	13.34	8.26	4.26	2.72	1.98	1.56	1.17	1.03
EWMA _{PRY}		500.03	56.62	17.52	8.93	5.54	2.94	1.95	1.48	1.23	1.03	1
EWMA _{PRX}		499.98	79.29	24.13	12.07	7.46	3.86	2.47	1.81	1.44	1.1	1.01
EWMA _{PMY}		500.43	53.53	16.59	8.4	5.26	2.78	1.86	1.42	1.19	1.03	1
EWMA _{PMX}		500.01	76.16	23.56	11.82	7.31	3.77	2.43	1.78	1.42	1.1	1.01
Classical EWMA	-0.50	500.65	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		499.23	63.57	19.36	9.75	6.06	3.16	2.08	1.55	1.27	1.04	1
EWMA _{PRY}		500.1	45.31	13.85	7.02	4.39	2.36	1.61	1.26	1.09	1.01	1
EWMA _{PRX}		500.45	63.32	19.15	9.56	5.92	3.09	2.02	1.51	1.24	1.03	1
EWMA _{PMY}		500.32	42.99	13.15	6.7	4.2	2.27	1.55	1.23	1.08	1	1
EWMA _{PMX}		500.25	61.45	18.78	9.44	5.82	3.06	2	1.5	1.23	1.03	1
Classical EWMA	-0.75	500.32	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		500.37	38.7	11.81	5.97	3.75	2.05	1.41	1.14	1.03	1	1
EWMA _{PRY}		499.98	27.78	8.57	4.4	2.81	1.6	1.18	1.04	1	1	1
EWMA _{PRX}		500.01	38.47	11.76	5.95	3.74	2.04	1.41	1.13	1.03	1	1
EWMA _{PMY}		500	27.11	8.24	4.24	2.71	1.56	1.16	1.03	1	1	1
EWMA _{PMX}		499.96	38.87	11.76	5.97	3.73	2.03	1.41	1.13	1.03	1	1
Classical EWMA	-0.95	500.1	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		500.04	15.97	4.99	2.62	1.72	1.11	1.01	1	1	1	1
EWMA _{PRY}		500	9.82	3.15	1.75	1.24	1.01	1	1	1	1	1
EWMA _{PRX}		499.96	12.72	3.98	2.14	1.46	1.04	1	1	1	1	1
EWMA _{PMY}		500.01	9.07	2.95	1.65	1.2	1	1	1	1	1	1
EWMA _{PMX}		499.95	12.31	3.88	2.09	1.44	1.03	1	1	1	1	1

is used. In all cases, the out-of-control ARL values approach 1 as shift increases.

The overall ARL performance of the modified control charts for different values of n and ρ_{XY} are studied using the extra quadratic loss (EQL) and the relative average run length (RARL) (Abujiya, Mu'azu Ramat and Farouk, Abbas Umar and Lee, Muhammad Hisyam and Mohamad, 2013; Wu et al., 2009). The EQL and RARL measure the efficiency of the modified charts over all values of the shifts considered. EQL is the weighted average of the ARL over the entire shifts of the charting structure, and makes use of the squares of the shifts (*i.e.*, δ^2) as the weights. The AEQL is given as:

$$EQL = \frac{1}{\delta_{max} - \delta_{min}} \int_{\delta_{min}}^{\delta_{max}} \delta^2 ARL(\delta) d(\delta). \quad (7.48)$$

Table 7.3: The ARL values of the modified EWMA control charts along with the ARL values of classical EWMA control chart when $C_X > C_Y$, and $n = 5$ for $\rho_{XY} = -0.25, -0.5, -0.75$ and -0.95 .

Method	ρ_{XY}	δ										
		0.25	0.25	0.5	0.75	1	1.5	2	2.5	3	4	5
Classical EWMA	-0.25	500	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		499.98	142.33	47.88	24.46	14.44	7.35	4.6	3.25	2.48	1.68	1.32
EWMA _{PRY}		500.01	130.56	42.53	21.13	13.03	6.62	4.15	2.97	2.27	1.57	1.25
EWMA _{PRX}		500.1	105.22	32.25	16.14	9.9	5.05	3.19	2.29	1.79	1.28	1.08
EWMA _{PMY}		500	127.92	41.56	20.68	12.69	6.48	4.07	2.9	2.23	1.55	1.24
EWMA _{PMX}		499.98	95.68	29.78	14.98	9.24	4.73	3.02	2.18	1.7	1.23	1.06
Classical EWMA	-0.50	500	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		500.01	112.98	35.08	17.45	10.71	5.43	3.45	2.47	1.92	1.37	1.13
EWMA _{PRY}		499.98	106.03	33.2	16.63	10.24	5.25	3.34	2.4	1.87	1.35	1.12
EWMA _{PRX}		500.01	77.95	24.36	12.26	7.54	3.91	2.51	1.85	1.47	1.12	1.02
EWMA _{PMY}		500.07	106.27	33.4	16.6	10.22	5.25	3.34	2.39	1.87	1.34	1.12
EWMA _{PMX}		499.97	77.94	23.68	11.83	7.27	3.75	2.42	1.78	1.42	1.1	1.01
Classical EWMA	-0.75	500	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		500.01	68.87	20.99	10.54	6.51	3.41	2.23	1.66	1.35	1.09	1.02
EWMA _{PRY}		500	69.98	21.15	10.55	6.5	3.4	2.22	1.66	1.35	1.08	1.01
EWMA _{PRX}		500.01	49.95	15.15	7.62	4.76	2.54	1.71	1.32	1.12	1.01	1
EWMA _{PMY}		500	67.31	20.57	10.38	6.43	3.37	2.2	1.65	1.34	1.08	1.01
EWMA _{PMX}		500	47.85	14.51	7.32	4.56	2.44	1.65	1.28	1.1	1.01	1
Classical EWMA	-0.95	500.56	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
EWMA _P		499.97	28.56	8.68	4.43	2.84	1.64	1.23	1.08	1.02	1	1
EWMA _{PRY}		500	22.22	6.89	3.59	2.33	1.39	1.1	1.02	1	1	1
EWMA _{PRX}		500.06	17.16	5.41	2.85	1.88	1.2	1.03	1	1	1	1
EWMA _{PMY}		500	21.7	6.69	3.47	2.27	1.36	1.09	1.01	1	1	1
EWMA _{PMX}		500	16.11	5	2.66	1.77	1.15	1.02	1	1	1	1

The chart with a smaller EQL is said to be superior than its counterpart chart (Wu et al., 2009).

The RARL measures the overall performance of a chart on a benchmark chart and is given as:

$$RARL = \frac{1}{\delta_{max} - \delta_{min}} \int_{\delta_{min}}^{\delta_{max}} \frac{ARL(\delta)}{ARL_{benchmark}(\delta)} d(\delta). \quad (7.49)$$

We use the chart from the classical EWMA chart as the benchmark chart, and $ARL_{benchmark}(\delta)$ is the ARL of the benchmark chart with shift δ . The chart with a RARL smaller than 1 is said to be superior to the benchmark chart (Abujiya, Mu'azu Ramat and Farouk, Abbas Umar and Lee, Muhammad Hisyam and Mohamad, 2013; Abbasi et al., 2015). The EQL and RARL are solved using numerical integration,

and δ_{max} and δ_{min} are the maximum and minimum shifts, respectively.

The results of the overall performance measures in Tables 7.4 and 7.5 show that the integration of an auxiliary variable, in the form of a product estimator, to the plotting statistic of the classical EWMA control chart, improves the overall performance of the classical EWMA control chart for values of the correlation within -0.95 to $-0.5C_X/C_Y$.

For any value of n and values of ρ_{XY} in this range (within -0.95 to $-0.5C_X/C_Y$), the modified control charts are more efficient than the classical EWMA control chart in detecting out-of-control signals. These efficiencies of the proposed charts over the classical chart are more pronounced as the value of the correlation coefficient between Y and X tends to -0.95 . For any values of ρ_{XY} in the range, an increase in the value of n leads to an improvement in the out-of-control detection ability of the proposed control charts, as well as the $EWMA_P$ control chart. For the $EWMA_P$ control chart, increases in the value of n do not substantially reduce the out-of-control ARL of the chart.

7.7 Illustrative examples

We provide illustrative examples of the proposed charts to show their application in real situations, using both simulated and real data sets. For the simulated example, we followed the approach in Abbas et al. (2014a), and generated a data set from a bivariate population. For the real data example, we used the nonisothermal continuous stirred tank chemical reactor model (CSTR), where the data set was originally given by Marlin (2000), and has been widely used as benchmark in fault detection and diagnosis (Abbasi and Riaz, 2013; Shi et al., 2013; Yoon and Macgregor, 2001). Specifically, our knowledge of the usage of the CSTR data set by Riaz (2015) prompted us to use the data for the real data validation.

Table 7.4: The EQL of the modified EWMA control charts along with the EQL of the classical EWMA control chart for different choices of the size of C_X and C_Y , different values of n and ρ_{XY} .

CV	n	ρ_{XY}	EWMA	EWMA _P	EWMA _{PRY}	EWMA _{PRX}	EWMA _{PMY}	EWMA _{PMX}
$C_X < C_Y$	5	-0.25	9.69	10.65	8.21	9.84	7.96	9.70
		-0.50	9.69	8.60	7.26	8.46	7.09	8.39
		-0.75	9.69	6.74	6.12	6.73	6.06	6.73
		-0.95	9.69	5.53	5.32	5.41	5.30	5.40
	10	-0.25	9.69	10.63	7.70	9.66	7.48	9.47
		-0.50	9.69	8.58	6.91	8.42	6.78	8.38
		-0.75	9.69	6.73	5.98	6.72	5.93	6.73
		-0.95	9.69	5.53	5.27	5.38	5.26	5.37
	15	-0.25	9.69	10.62	7.46	9.50	7.30	9.55
		-0.50	9.69	8.58	6.80	8.37	6.69	8.35
		-0.75	9.69	6.73	5.92	6.72	5.88	6.73
		-0.95	9.69	5.53	5.26	5.38	5.24	5.37
$C_X = C_Y$	5	-0.25	9.69	12.67	10.64	10.61	10.46	10.29
		-0.50	9.69	9.74	8.97	8.80	8.78	8.70
		-0.75	9.69	7.07	6.88	6.85	6.85	6.82
		-0.95	9.69	5.37	5.35	5.35	5.35	5.35
	10	-0.25	9.69	12.66	10.07	10.00	9.90	9.86
		-0.50	9.69	9.74	8.68	8.59	8.57	8.52
		-0.75	9.69	7.06	6.80	6.79	6.81	6.78
		-0.95	9.69	5.36	5.35	5.35	5.35	5.35
	15	-0.25	9.69	12.65	9.85	9.84	9.74	9.68
		-0.50	9.69	9.73	8.59	8.47	8.51	8.48
		-0.75	9.69	7.06	6.80	6.77	6.79	6.76
		-0.95	9.69	5.36	5.35	5.35	5.35	5.35
$C_X > C_Y$	5	-0.25	9.69	16.99	15.57	12.26	15.28	11.60
		-0.50	9.69	13.00	12.77	9.98	12.74	9.71
		-0.75	9.69	9.05	9.11	7.54	9.03	7.38
		-0.95	9.69	6.17	5.76	5.49	5.74	5.45
	10	-0.25	9.69	16.97	15.13	10.87	14.92	10.46
		-0.50	9.69	13.00	12.58	9.31	12.67	8.99
		-0.75	9.69	9.05	9.07	7.20	8.98	7.04
		-0.95	9.69	6.17	5.76	5.49	5.74	5.45
	15	-0.25	9.69	16.96	15.00	10.51	14.73	10.07
		-0.50	9.69	12.99	12.55	9.01	12.49	8.70
		-0.75	9.69	9.01	9.02	7.04	8.96	6.98
		-0.95	9.69	6.15	5.74	5.45	5.72	5.42

Table 7.5: The RARL of the modified EWMA control charts along with the RARL of the classical EWMA control chart for different choices of the size of C_X and C_Y , different values of n and ρ_{XY} .

CV	n	ρ_{XY}	EWMA	EWMA _P	EWMA _{PRY}	EWMA _{PRX}	EWMA _{PMY}	EWMA _{PMX}
$C_X < C_Y$	5	-0.25	1	1.11	0.82	1.02	0.79	1.00
		-0.50	1	0.87	0.70	0.85	0.68	0.84
		-0.75	1	0.64	0.55	0.63	0.54	0.63
		-0.95	1	0.47	0.43	0.45	0.43	0.44
	10	-0.25	1	1.10	0.76	1.00	0.73	0.97
		-0.50	1	0.87	0.66	0.85	0.64	0.84
		-0.75	1	0.63	0.53	0.63	0.52	0.64
		-0.95	1	0.47	0.42	0.44	0.42	0.44
	15	-0.25	1	1.10	0.73	0.98	0.71	0.97
		-0.50	1	0.87	0.64	0.84	0.63	0.84
		-0.75	1	0.63	0.52	0.63	0.52	0.63
		-0.95	1	0.46	0.42	0.44	0.42	0.44
$C_X = C_Y$	5	-0.25	1	1.34	1.10	1.11	1.08	1.07
		-0.50	1	1.01	0.91	0.89	0.89	0.88
		-0.75	1	0.68	0.65	0.65	0.65	0.65
		-0.95	1	0.44	0.44	0.44	0.44	0.44
	10	-0.25	1	1.33	1.04	1.03	1.02	1.02
		-0.50	1	1.00	0.88	0.87	0.86	0.86
		-0.75	1	0.68	0.64	0.64	0.64	0.64
		-0.95	1	0.44	0.44	0.44	0.44	0.44
	15	-0.25	1	1.33	1.01	1.02	1.00	1.00
		-0.50	1	1.00	0.87	0.85	0.86	0.86
		-0.75	1	0.68	0.64	0.64	0.64	0.64
		-0.95	1	0.44	0.44	0.44	0.44	0.44
$C_X > C_Y$	5	-0.25	1	1.82	1.65	1.29	1.61	1.21
		-0.50	1	1.39	1.34	1.03	1.34	1.00
		-0.75	1	0.93	0.93	0.74	0.91	0.72
		-0.95	1	0.56	0.50	0.46	0.50	0.45
	10	-0.25	1	1.80	1.60	1.13	1.58	1.09
		-0.50	1	1.37	1.32	0.96	1.34	0.92
		-0.75	1	0.92	0.92	0.70	0.91	0.67
		-0.95	1	0.56	0.50	0.46	0.50	0.45
	15	-0.25	1	1.78	1.59	1.10	1.56	1.04
		-0.50	1	1.37	1.31	0.92	1.31	0.88
		-0.75	1	0.92	0.92	0.67	0.91	0.67
		-0.95	1	0.55	0.50	0.45	0.49	0.45

7.7.1 Simulated dataset

We generate 20 data sets each of $n^2 = 25$ from $(Y, X) \sim N_2(\mu_1 = \mu_Y + \delta * (\sigma_Y / \sqrt{n}), \mu_X = 50, \sigma_Y = 14, \sigma_X = 5, \rho_{XY} = -0.75)$, where $\mu_Y = 100$. From each data set of size $n^2 = 25$, we draw a sample of

size $n = 5$ using different sampling schemes considered in the study. The smoothing parameter used is $\lambda = 0.05$, $L = 2.639$ for the classical chart, and $L_c = 2.646$ for the proposed charts. These values are chosen to fix the in-control ARL of the charts at 500. The results are given in Figures 7.4 to 7.9.

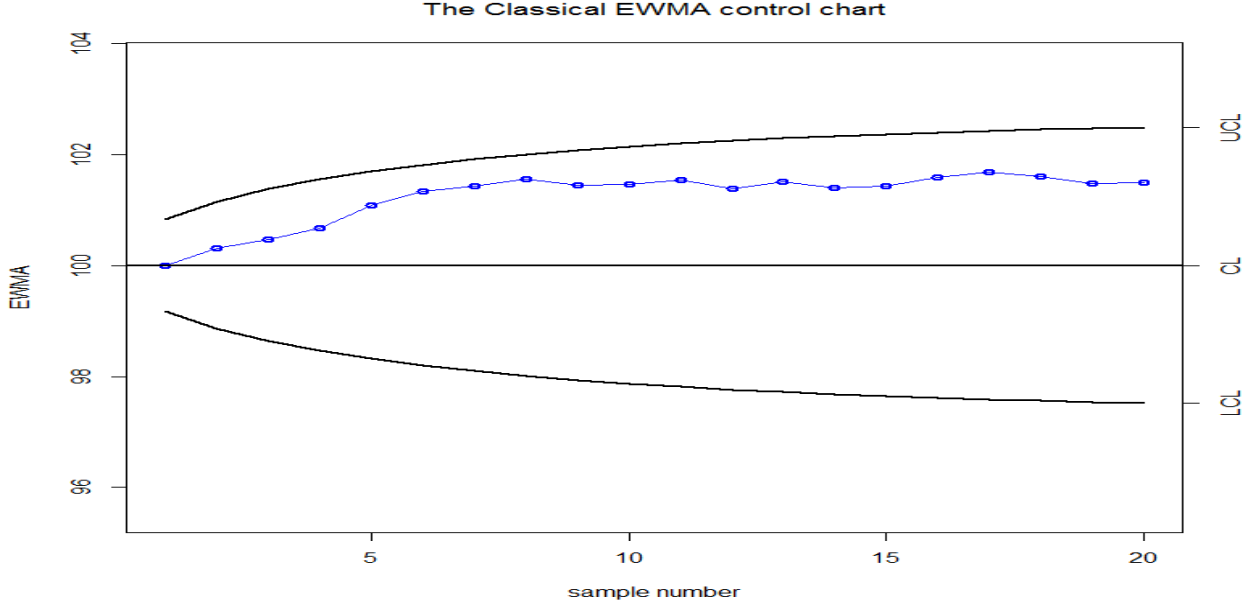


Figure 7.4: The classical EWMA control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

As shown in Figures 7.4 to 7.9, the classical EWMA control chart fails to detect the shift in the process mean. However, all of the proposed charts detect out-of-control signals in the process mean. The $EWMA_P$ gives the first out-of-control signal after the 17th sample, while the $EWMA_{PRX}$ gives the first out-of-control signal after the 14th sample. The proposed $EWMA_{PMX}$ detects the first out-of-control signal after the 10th sample; however, the chart later gives an in-control condition after the 15th sample. Both the $EWMA_{PRY}$ and $EWMA_{PMY}$ detect the first shift after the 11th and 10th samples. Since $C_X < C_Y$, the proposed charts based on MRSS and RSS schemes on Y are more efficient than the proposed charts based on MRSS and RSS on X in detecting the shift.

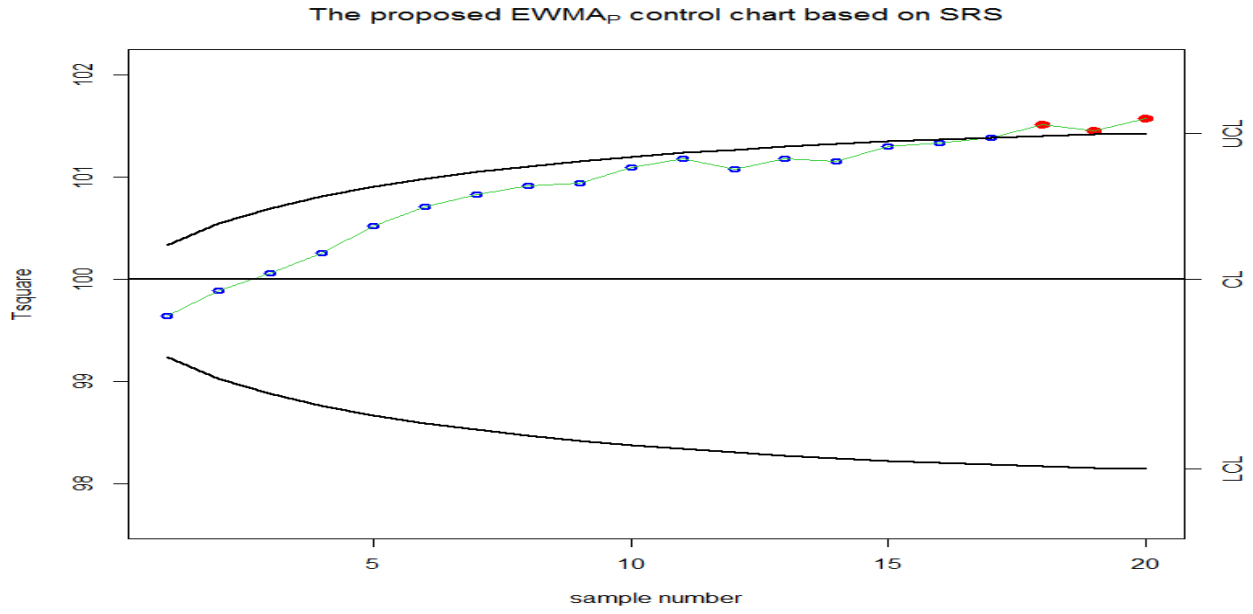


Figure 7.5: The proposed $EWMA_P$ control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

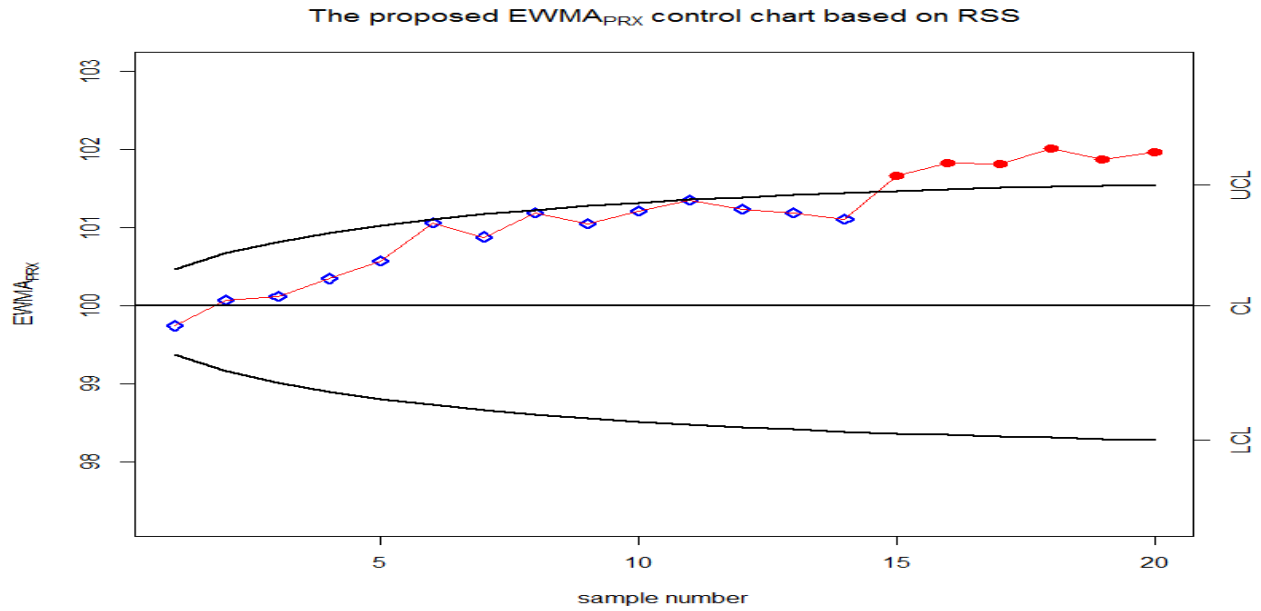


Figure 7.6: The proposed $EWMA_{PRX}$ control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

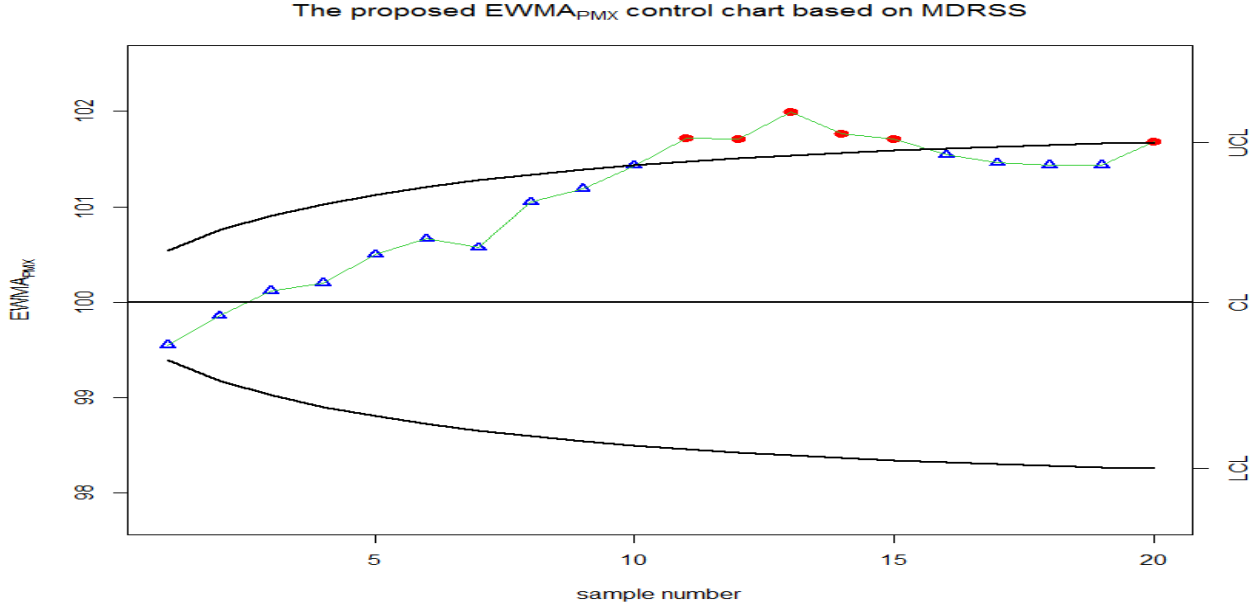


Figure 7.7: The proposed $EWMA_{PMX}$ control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

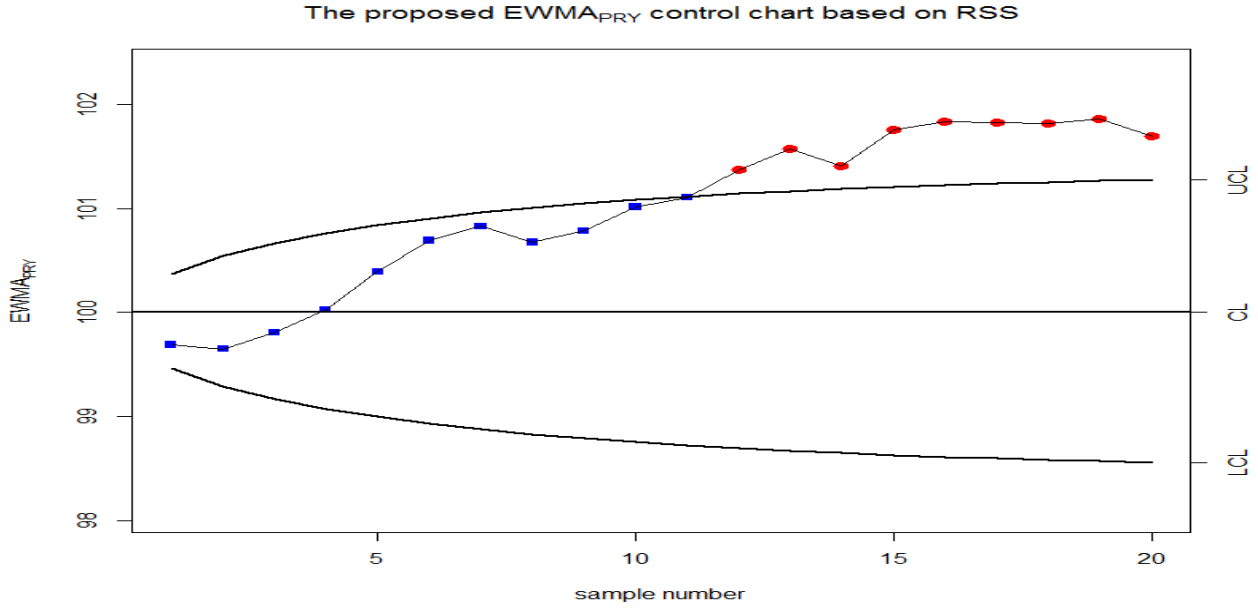


Figure 7.8: The proposed $EWMA_{PRY}$ control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

7.7.2 Real dataset

The CSTR process comprises nine process variables, among which we have used the outlet concentration (CA in $kmole/m^3$) as the process variable Y . Many authors have also used this variable as the process

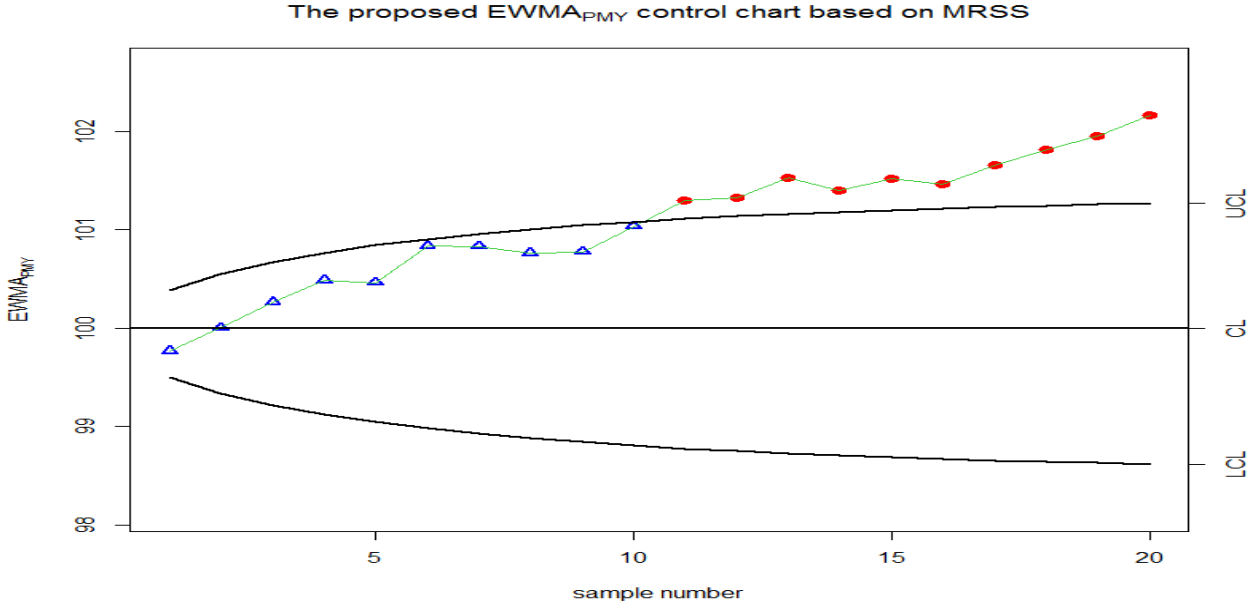


Figure 7.9: The proposed $EWMA_{PMY}$ control chart with time varying limits of the simulated data from the bivariate normal distribution with $\mu_Y = 100$, $\mu_X = 50$, $\sigma_Y = 14$, $\sigma_X = 5$, $\rho_{XY} = -0.75$ and $\delta = 0.5$, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

variable, for example, see Abbasi and Riaz (2013), and flow rate of reactant (FA in m^3/min) as X. The data set originally contains 1024 samples, collected on a sampling interval of half minute. Details of other variables may be found in Yoon and Macgregor (2001) and Shi et al. (2013).

The first 512 values are obtained when the process was in an in-control state, and are used as the Phase I data set. The parameters of the Phase I samples are $\mu_Y = 0.8033647$, $\sigma_Y = 0.05699951$, $\sigma_X = 0.007082823$, $\mu_X = 0.0988881$ and the correlation between the variables is $\rho_{XY} = -0.5860537$. The coefficient of variations of the Phase I samples are $C_X = 0.07162462$, and $C_Y = 0.07095098$. In this case, we have $C_X > C_Y$. These parameters are used to estimate the product estimator from a sample of size $n = 5$, using different sampling schemes considered in this study, and the parameters are estimated from 100,000 simulations of the product estimators.

For the Phase II monitoring, we used the first 500 of the last 512 samples as Phase II. Our initial investigation showed that a shift of size $\delta = \frac{|\mu_Y - \mu_1|}{\sigma_Y/\sqrt{n}} = 0.2744592$, was introduced in Phase II, where μ_1 is the Phase II mean, and $n = 5$ is the selected sample size used in calculating the product estimator

at each time point in Phase II. Thus, the 500 Phase II data set is grouped into 20 data sets each of $n^2 = 25$. For the SRS, we select $n = 5$ random samples of the study variable from each of these 20 data sets. Hence, for an actual value of the study variable, we select the corresponding value of the auxiliary variable. However, for both the MRSS and RSS, we divide each of the 20 data sets randomly into five sets of size 5. By visual inspection, we assign rank to each set with respect to the auxiliary variable (since $C_X > C_Y$), and draw a sample of size $n = 5$ samples of the auxiliary variable, using each of the schemes. For an actual value of the auxiliary variable obtained from the MRSS and RSS, we select the corresponding value of the study variable. At each time point in Phase II, these $n = 5$ samples are used in calculating the product estimator used in the modified EWMA statistic for each chart. The results are given in Figures 7.10-7.15.

As shown in Figures 7.10 to 7.15, the classical EWMA, $EWMA_P$, $EMWA_{PRY}$ and $EWMA_{PMY}$ control charts fail to detect the shifts in the process mean. However, both the $EMWA_{PRX}$ and $EWMA_{PMX}$ detect shifts. The $EWMA_{PRX}$ gives the first out-of-control signal after the 15th sample, while the $EWMA_{PMX}$ gives the first out-of-control signal after the 14th sample. The results of the real life example support the results in Tables 7.4 and 7.5. When $C_X > C_Y$, the performance of the proposed charts based on RSS and MRSS on X are more efficient in detecting the out-of-control signals than the classical EWMA control chart, for values of $\rho_{XY} < -0.5C_X/C_Y$.

7.8 General conclusions

In this study, modifications of the classical EWMA control chart are proposed to suit the situation where the process variable is negatively correlated with the auxiliary information. The control charts for different sampling schemes, that is, simple random sampling, ranked set sampling, and median ranked set sampling, are developed. Also, recommendations are given on which of the variables the ranked set sampling and

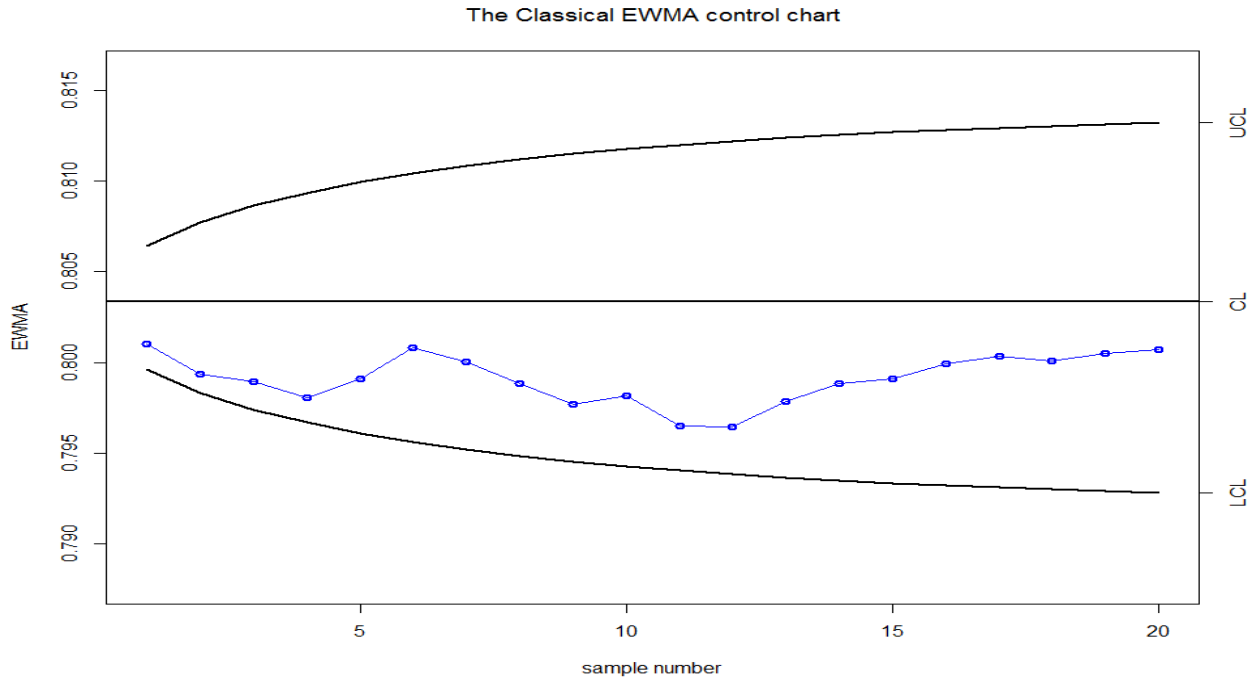


Figure 7.10: The classical EWMA control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

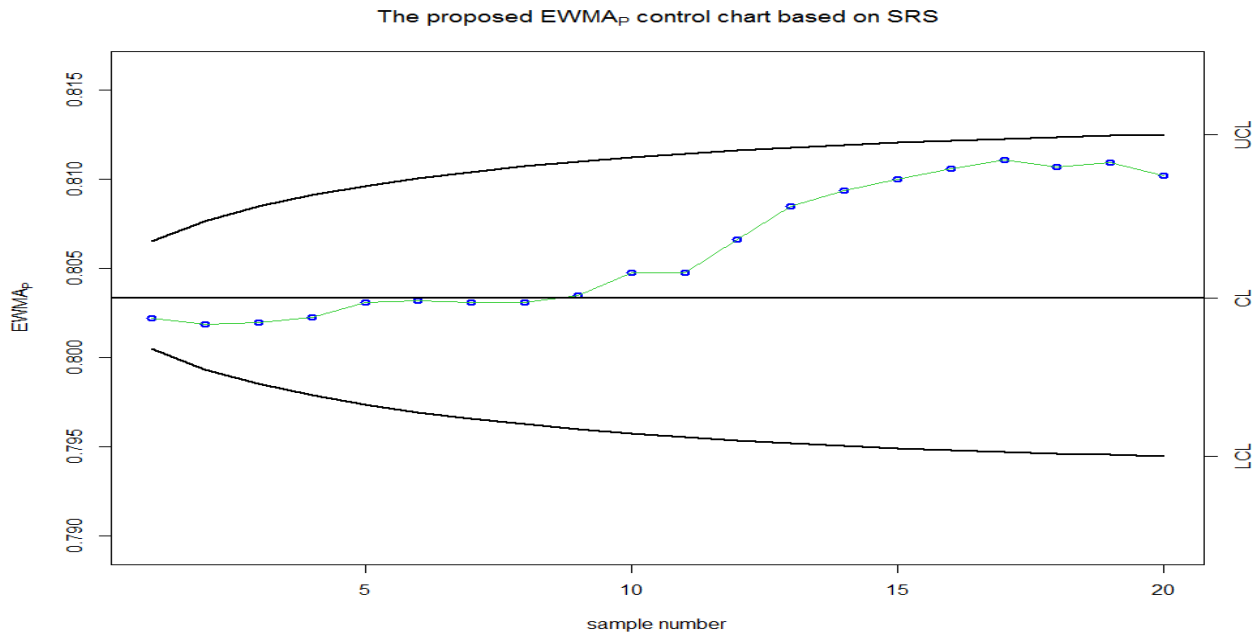


Figure 7.11: The proposed $EWMA_P$ control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

the median ranked set sampling should be applied. We showed, through simulation, that the proposed charts are more efficient than the classical EWMA chart in detecting shifts in the location parameter of

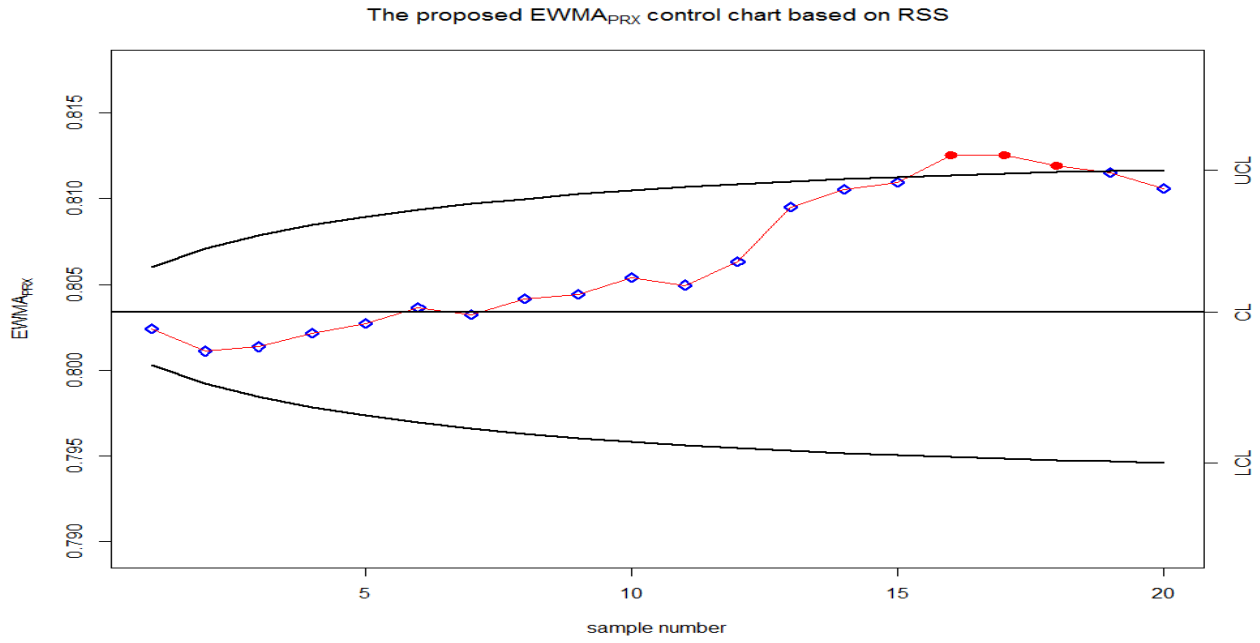


Figure 7.12: The proposed $EWMA_{PRX}$ control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

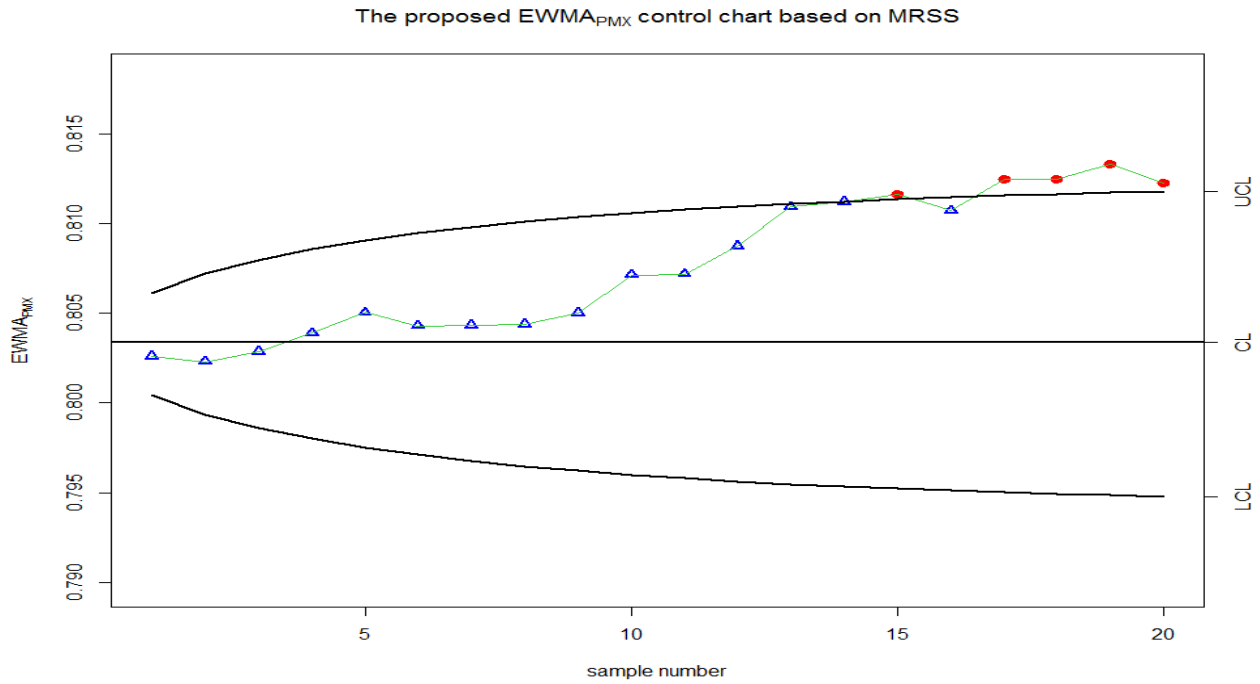


Figure 7.13: The proposed $EWMA_{PMX}$ control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

a production process where the process variable is negatively correlated with the auxiliary variable, for some range of correlation between the variables. The efficiency of the proposed charts over the classical

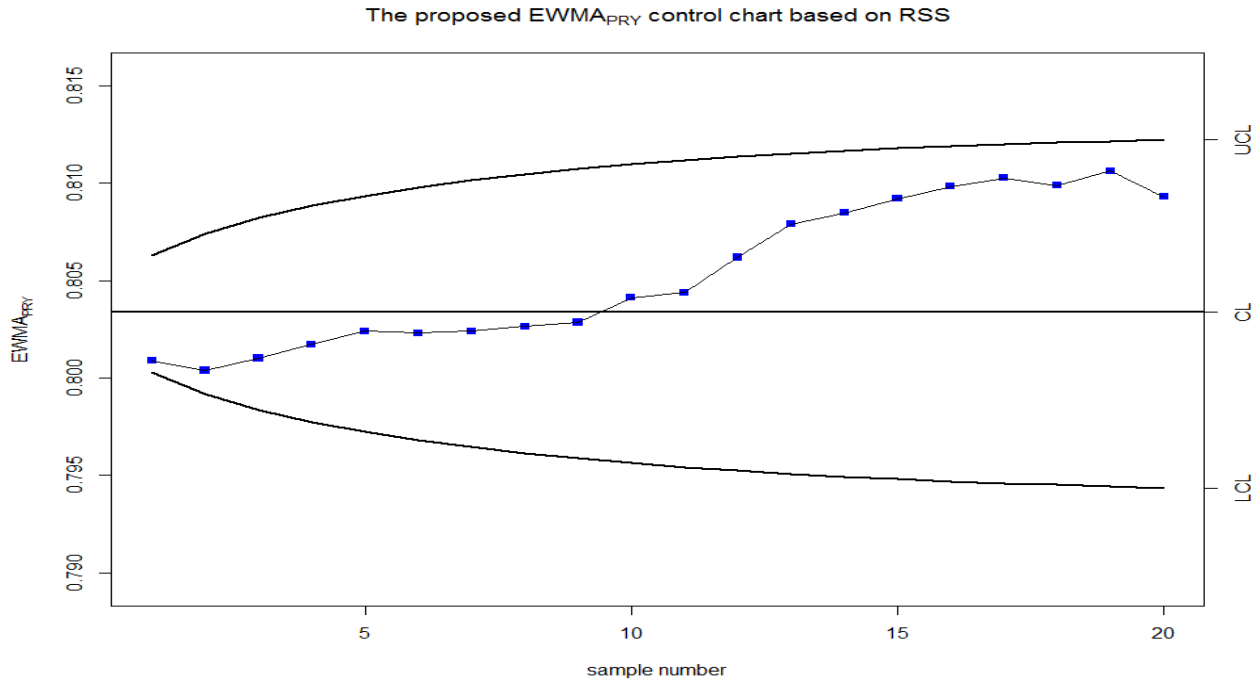


Figure 7.14: The proposed $EWMA_{PRY}$ control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

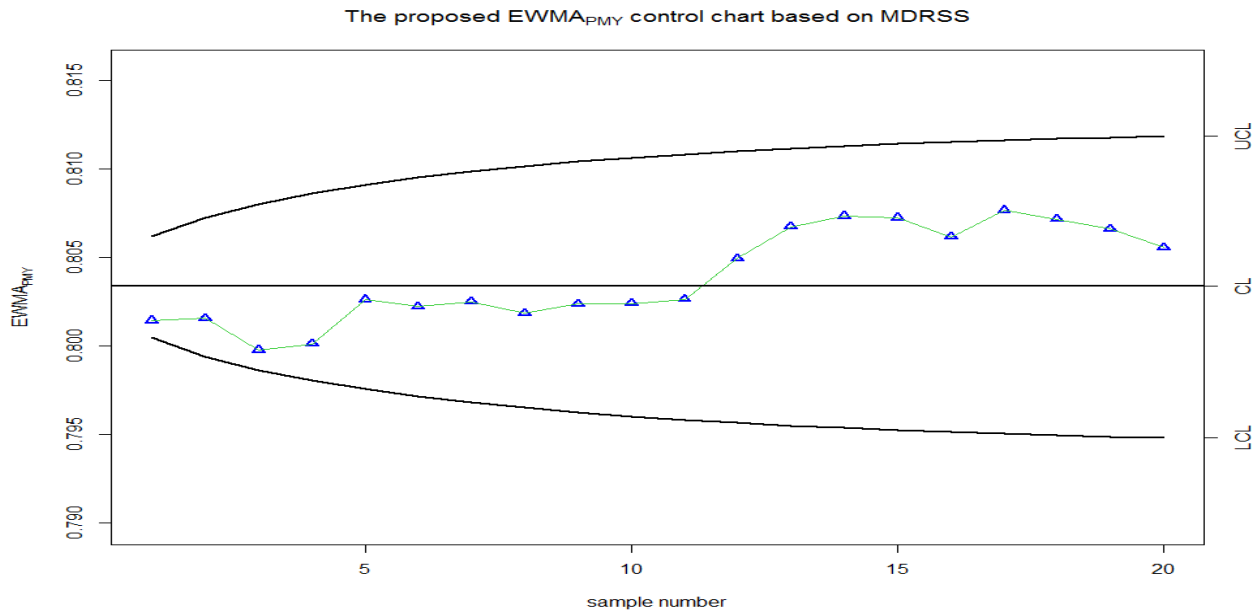


Figure 7.15: The proposed $EWMA_{PMY}$ control chart with time varying limits of the real life example, where the parameters of the chart are $\lambda = 0.05$ and $L_c = 2.646$ with $ARL_0 = 500$.

EWMA control chart is more pronounced when there is a large negative correlation between the variables.

Furthermore, the proposed charts are very efficient in detecting small shifts. Our simulation results show

that the proposed charts based on the median ranked set sampling, followed by the ranked set sampling are more efficient than the proposed charts based on simple random sampling. The performance of the proposed control charts is also illustrated with examples.

In Sanusi et al. (2017b) (see Appendix A.1 for reference), we extend the research, here, to the case of positive correlations between the variables and proposed an EWMA-type chart, called an M_r EWMA, uses the auxiliary variable in the form of ratio estimator. There, the auxiliary variable must be strongly positively related to the study variable for the proposed M_r EWMA chart to perform better than its existing counterparts. Except for the case of weak correlation value, it was shown that the proposed M_r EWMA chart outperforms the classical EWMA chart in detecting small to moderate shifts in the location parameter of a control process. Also, the proposed M_r EWMA chart performs better than the mixed EWMA-CUSUM chart when $\rho_{XY} \geq 0.50$.

Chapter 8

EWMA control chart for monitoring the mean of a process that is correlated with an auxiliary variable under some ranked sampling scheme

”This is the peer reviewed version of the following article: “Adegoke, N. A., Abbasi, S. A., Dawod, A. B., & Pawley, M. D. Enhancing the performance of the EWMA control chart for monitoring the process mean using auxiliary information. *Quality and Reliability Engineering International*”, which has been published in final form at <https://doi.org/10.1002/qre.2436>. This article may be used for non-commercial purposes in accordance with Wiley Terms and Conditions for Self-Archiving.”

When using control charts to monitor manufacturing processes, the exponentially weighted moving average (EWMA) control chart is useful for detecting persistent shifts in the process parameter. This paper

proposes enhancements to the applications of the EWMA control chart for those scenarios where the exact measurement of process units is difficult and expensive, but the visual ordering of the units can be done easily. The proposed charts use an auxiliary variable that is correlated with the process variable to provide efficient monitoring of shifts in the process mean, and are formulated based on ranked set sampling (RSS) and median ranked set sampling schemes (MRSS). Simulation results showed that the proposed charting schemes are more efficient in detecting a shift in the process mean than the classical EWMA control chart and its modification. An example is provided to show the application of the proposed charts using a simulated benchmark process: the continuous stirred tank reactor (CSTR).

8.1 Introduction

Control charts are classified into memoryless-type and memory-type control charts (Ajadi et al., 2016). The Shewhart control chart is a memoryless-type control chart that utilizes only the current information without referring to the previous information in the monitoring process (Shewhart, 1931); it is useful for detecting large shifts in the process parameter. The EWMA control chart (Roberts, 1959), and the cumulative sum (CUSUM) control chart (Page, 1954), are the two most commonly adopted memory-type control charts. They use both the previous and current information in the monitoring process and are well suited for detecting small to moderate shifts in process parameters.

Control charts have been used in situations where the process variables are observed with another variable. Many researchers have investigated the importance of using this information to enhance the performance of the traditional control charts, and several forms of the models that incorporate both the process and auxiliary variables have been proposed in SPC literature (see Shabbir and Awan (2016); Abbasi and Riaz (2013); Haq et al. (2016); Riaz et al. (2013)). One such practical model is based on the regression estimation procedures (Cochran, 1977). For example, Riaz (2008a,b) utilized regression-type

estimators to monitor the location and dispersion parameters, respectively. Their results outperformed the Shewhart-type control charts used for the same purposes, i.e., R , S and S^2 charts for monitoring the variability, and \bar{X} chart for monitoring the location parameter.

More recently, Abbas et al. (2014a) proposed an EWMA-type control chart (called the M_X EWMA chart) in the form of a regression estimator for monitoring shifts in the location parameter. In their work, both the process and auxiliary variables were obtained using a simple random sampling (SRS) scheme. The M_X EWMA chart used the extra information to provide an efficient estimator of the location parameter. Their proposed chart out-performed the classical EWMA control chart for monitoring shifts in the location parameter when there was a large correlation between the variables.

The performance of control charts in identifying assignable shifts in the production and manufacturing process depends on the sampling scheme used in their development (Abbas et al., 2014a). Developing control charts using more efficient sampling schemes, such as ranked set sampling (RSS), or median ranked set sampling (MRSS), can enhance productivity and reduce production costs (Haq et al., 2016). In this paper, we propose a EWMA-type control charts based on a regression estimator, for monitoring shifts in the process mean using an auxiliary variable that is correlated with the process variable. The proposed charts are formulated using RSS and MRSS, and use the extra information along with the process variable in the monitoring process. Practitioners will be able to apply the proposed EWMA control charts in those situations where exact measurements of units are difficult and expensive, but the visual ordering of units can be done easily. The performance of the proposed charts is examined using average run length (ARL) properties; ARL is the average number of plotted statistics until a shift is detected.

The rest of the chapter is organized as follows: Section 8.2 reviews the RSS and MRSS schemes. Section 8.3 describes the proposed charts. A simulation study is discussed in Section 8.4. Section 8.5 discusses the main finding of the simulation study. Section 8.6 presents an illustrative example. Conclusions and recommendations are presented in Section 8.7.

8.2 The ranked set sampling and median ranked set sampling schemes

Let Y_1, Y_2, \dots, Y_n be a random sample of size n , from probability density function $f(y)$. The SRS estimator of the mean of the process variable, Y , is given as $\bar{Y} = \frac{1}{n} \sum_{i=1}^n Y_i$, with standard error $sd(\bar{Y}) = \frac{\sigma_Y}{\sqrt{n}}$. Let $Y_{11}, Y_{12}, \dots, Y_{1n}, Y_{21}, Y_{22}, \dots, Y_{2n}, \dots, Y_{n1}, Y_{n2}, \dots, Y_{nn}$ be n independent SRS each of size n , and $Y_{(i(1:n))}, Y_{(i(2:n))}, Y_{(i(3:n))}, \dots, Y_{(i(n:n))}$ represent the order statistics of the i th sample. The RSS starts with n^2 units and divides the units into n sets, each of size n . Then, rank the samples in each set with respect to the process variable. Pick the smallest value from the first set, the second-smallest from the second set, etc, until we reach the n th set. This results in a sample of size n which represents the RSS data (McIntyre, 1952). The process may be cycled m ways until $k = nm$ units are chosen. The nm units constitute the RSS units.

The measured RSS units are denoted by $Y_{1(1:n)}, Y_{2(2:n)}, Y_{3(3:n)}, \dots, Y_{n(n:n)}$. Let $g_{(i:n)}(y)$ be the probability density function of the i th order statistics $Y_{(i:n)}, i = 1, 2, 3, \dots, n$, from a random sample of size n . It can be shown that:

$$g_{(i:n)}(y) = n \binom{n-1}{i-1} \{F(y)\}^{i-1} \{1-F(y)\}^{n-i} f(y) \quad -\infty < y < +\infty \quad (8.1)$$

where $F(y)$ is the cumulative distribution of Y (David, H and Nagaraja, 2003; Haq et al., 2013). The mean and variance of $Y_{(i:n)}$ are given by:

$$\mu_{(i:n)} = \int_{-\infty}^{+\infty} yg_{(i:n)}(y)dy \quad \text{and} \quad \sigma_{(i:n)}^2 = \int_{-\infty}^{+\infty} (y - \mu_{(i:n)})^2 g_{(i:n)}(y)dy \quad (8.2)$$

Following Takahasi and Wakimoto (1968), the mean and the variance of the RSS are given by::

$$\bar{Y}_{rss,k} = \frac{1}{n} \sum_{i=1}^n Y_{i\{i:n\}}, k = 1, 2, \dots, r \quad (8.3)$$

and

$$Var(\bar{Y}_{RSS}) = \frac{1}{n^2} \sum_{i=1}^n \sigma_{\{i:n\}}^2 = \frac{\sigma_Y^2}{n} - \frac{1}{n^2} \sum_{i=1}^n (\mu_{(i:n)} - \mu_Y)^2 \quad (8.4)$$

$Y_{i\{i:n\}}$ represents the order statistics of the i th sample, and μ_Y is the population mean of Y .

The MRSS scheme is a modified version of the RSS scheme (Muttalak, 1997), and can be summarized as follows: after ranking each set on the interest variable, we have two possibilities depending on whether the n is even or odd. If the sample size is odd, choose $(\frac{n+1}{2})$ th value of each ranked set. If the sample size is even, partition the units into two equal parts and select $[(\frac{n}{2})]$ th smallest ranked values from each set in the first half, and $(\frac{n+2}{2})$ th smallest ranked values from each set in the second half. This process may be cycled m times until $k = nm$ units are chosen. The nm units constitute the MRSS units. Analytically, the mean and variance of the MRSS at the k th cycle for both the even and odd sample size n are given as:

$$\bar{Y}_{MRSS}^{Even} = \frac{1}{k} \sum_{j=1}^m \left(\sum_{i=1}^{n/2} Y_{i\{(n/2):n\}j} + \sum_{i=(n/2)+1}^n Y_{i\{(n/2)+1:n\}j} \right) \quad \text{and} \quad \bar{Y}_{MRSS}^{Odd} = \frac{1}{k} \sum_{j=1}^m \sum_{i=1}^n Y_{i\{(n+1)/2:n\}j} \quad (8.5)$$

\bar{Y}_{MRSS}^{Even} and \bar{Y}_{MRSS}^{Odd} are unbiased estimates of μ_Y , for any symmetric distribution (David, H and Nagaraja, 2003; Haq et al., 2013). Their variances, respectively, are given as:

$$Var(\bar{Y}_{MRSS}^{Even}) = E(\bar{Y}_{MRSS}^{Even} - \mu_Y)^2 - \frac{1}{kn} \left(\sum_{i=1}^{n/2} \sigma_{Y\{(n/2):n\}}^2 + \sum_{i=(n/2)+1}^n \sigma_{Y\{(n/2)+1:n\}}^2 \right)$$

$$Var(\bar{Y}_{MRSS}^{Odd}) = E(\bar{Y}_{MRSS}^{Odd} - \mu_Y)^2 - \frac{1}{kn} \sum_{i=1}^n \sigma_{Y\{(n+1)/2:n\}}^2$$

The RSS and MRSS estimates of the mean and standard deviation in this section and other sections are obtained numerically. In all cases, we used $m=1$. We refer the reader to David, H and Nagaraja (2003), for detailed analytical theory on RSS and MRSS.

8.3 The proposed charts

Assume that Y_i and X_i are respectively, a process and auxiliary variable that jointly occur in pairs at each time i with correlation ρ_{XY} . We assume the joint distribution of Y and X can be modelled using a bivariate normal distribution. Following Adegoke et al. (2017), we examined the case of correctly applying the RSS and MRSS schemes on one variable at a particular time. We refer to the scenario where Y is correctly ranked while the ranking on X is done with error, as case I. In contrast, the case of correctly applying the ranking on X while the ranking on Y is done with error, is referred to as case II.

For case I, we define $(Y_{(i:n)k}, X_{([i:n]k)})$ to be the i th smallest observation of Y , associated with the corresponding value of X , obtained from the i th set in the k th cycle. The regression estimate of the population mean (μ_Y) under the RSS is given as follows:

$$S_Y = Y_{rss,k} + \hat{\beta}_{1,k}(\mu_X - \bar{X}) \quad (8.6)$$

where $\bar{Y}_{rss,k}$ is defined in Equation (8.3), and $\hat{\beta}_{1,k} = \frac{\sum_{i=1}^n (Y_{(i:n)k} - \bar{Y}_{rss,k})(X_{(i:n)k} - \bar{X})}{E(X_{(i:n)k} - \bar{X})^2}$. The mean and variance of S_Y are given as:

$$E(S_Y) = \mu_Y, \quad \sigma_{S_Y}^2 = \frac{\sigma_{\bar{X}mrss,k}^2}{n}(1 - \rho_{XY}^2) \quad (8.7)$$

Considering the proposed estimator S_Y in Equation (8.6), the plotting statistic for the proposed chart

based on RSS (hereafter S_XEWMA) at time i is given as:

$$W_i = \Phi S_{Yi} + (1 - \phi)W_{(i-1)}, \text{ where } W_0 = \mu_Y. \quad (8.8)$$

where $W_{(i-1)}$ represents past information with $W_0 = \mu_y$ and $\phi, (0 < \phi \leq 1)$ is the smoothing parameter which determines the rate at which past information is incorporated into the calculation of the EWMA statistic (Abbas et al., 2014a). The control limits of the S_XEWMA chart are given as:

$$\begin{aligned} LCL &= \mu_Y - L\sigma_{S_Y} \sqrt{\frac{\phi}{2-\phi}(1 - (1-\phi)^{2i})} \\ CL &= \mu_Y \\ UCL &= \mu_Y + L\sigma_{S_Y} \sqrt{\frac{\phi}{2-\phi}(1 - (1-\phi)^{2i})} \end{aligned} \quad (8.9)$$

Under the MRSS scheme, the regression estimate of the population mean (μ_Y) is given as:

$$T_Y = Y_{mrss,k} + \hat{\beta}_{2,k}(\mu_X - \bar{X}) \quad (8.10)$$

where $\bar{Y}_{mrss,k}$ is defined in Equation (8.5), and $\hat{\beta}_{2,k} = \frac{\sum_{i=1}^n (Y_{(i:n)k} - \bar{Y}_{mrss,k})(X_{(i:n)k} - \bar{X})}{E(X_{(i:n)k} - \bar{X})^2}$. The mean and variance of S_Y are given as:

$$E(T_Y) = \mu_Y, \quad \sigma_{T_Y}^2 = \frac{\sigma_{\bar{X}mrss,k}^2}{n}(1 - \rho_{XY}^2) \quad (8.11)$$

The plotting statistic for the proposed chart based on MRSS (hereafter T_XEWMA) at time i is given as:

$$P_i = \Phi T_{Yi} + (1 - \phi)T_{(i-1)}, \text{ where } T_0 = \mu_Y. \quad (8.12)$$

where $T_{(i-1)}$ represents past information with $T_0 = \mu_y$ and $\phi, (0 < \phi \leq 1)$ is the smoothing parameter

which determines the rate at which past information is incorporated into the calculation of the EWMA statistic (Abbas et al., 2014a). The control limits of the T_X EWMA chart are given as:

$$\begin{aligned} LCL &= \mu_Y - L\sigma_{T_Y} \sqrt{\frac{\phi}{2-\phi}(1-(1-\phi)^{2i})} \\ CL &= \mu_Y \\ UCL &= \mu_Y + L\sigma_{T_Y} \sqrt{\frac{\phi}{2-\phi}(1-(1-\phi)^{2i})} \end{aligned} \tag{8.13}$$

where L is a constant, selected arbitrarily to manipulate the width of the control limits, (the values of L and ϕ are chosen to fix the in-control ARL to the desired value) (Maravelakis et al., 2004; Abbasi et al., 2015). Analogous to the derivation defined for case I, the limits for case II can be obtained by defining $(Y_{[i:n]_k}, X_{(i:n)_k})$ to be i th smallest observation of X , associated with the corresponding value of Y , obtained from the i th set in the k th cycle.

We studied the efficiency of the regression estimators based on the different sampling schemes over the conventional estimator of Y for both case I and II. Specifically, we found the range of such that:

$$sd(\bar{Y}^*(\rho_{XY})) < sd(\bar{Y}) \tag{8.14}$$

where we have used $sd(\bar{Y}^*(\rho_{XY}))$ to represent the standard error of the regression estimator from a particular sampling scheme, i.e., the SRS, RSS, and MRSS, besides the conventional estimator (i.e., \bar{Y}), and $sd(\bar{Y})$ is the standard error of the conventional estimator. We generated $n \in \{5, 10\}$ bivariate normally distributed data using the parameters and simulation procedures (of Phase I) given in Section 8.3, and estimated the standard error of the estimator from each of the sampling schemes.

The simulation results in Figure 8.1 and Figure 8.2 show that the standard error of the regression estimator based on each of the sampling schemes is more efficient than the conventional estimator for all values of ρ_{XY} . For case I, we found both RSS and MRSS are more efficient than SRS when (see

Figure 8.1). However, for case II, neither RSS nor MRSS performed better than SRS (see Figure 8.2). The simulation results in Figure 8.1 affirm the results in Patil, GP and Sinha, AK and Taille (1993) for RSS, where it was also shown that the RSS regression estimator was considerably more efficient than the SRS regression estimator unless the correlation between the auxiliary variable and the process variable was greater than or equal to 0.85 (i.e., $|\rho_{XY}| \geq 0.85$). Figure 8.2 shows that when ranking is done by the auxiliary variable, the performance of RSS and SRS are comparable. We refer interested readers to Patil, GP and Sinha, AK and Taille (1993) for a comprehensive study of the efficiency of both cases I and II. Based on the preliminary investigation in Figure 8.1 and Figure 8.2, Section 4 describes a comprehensive study of ARL performance of the proposed charts from case I.

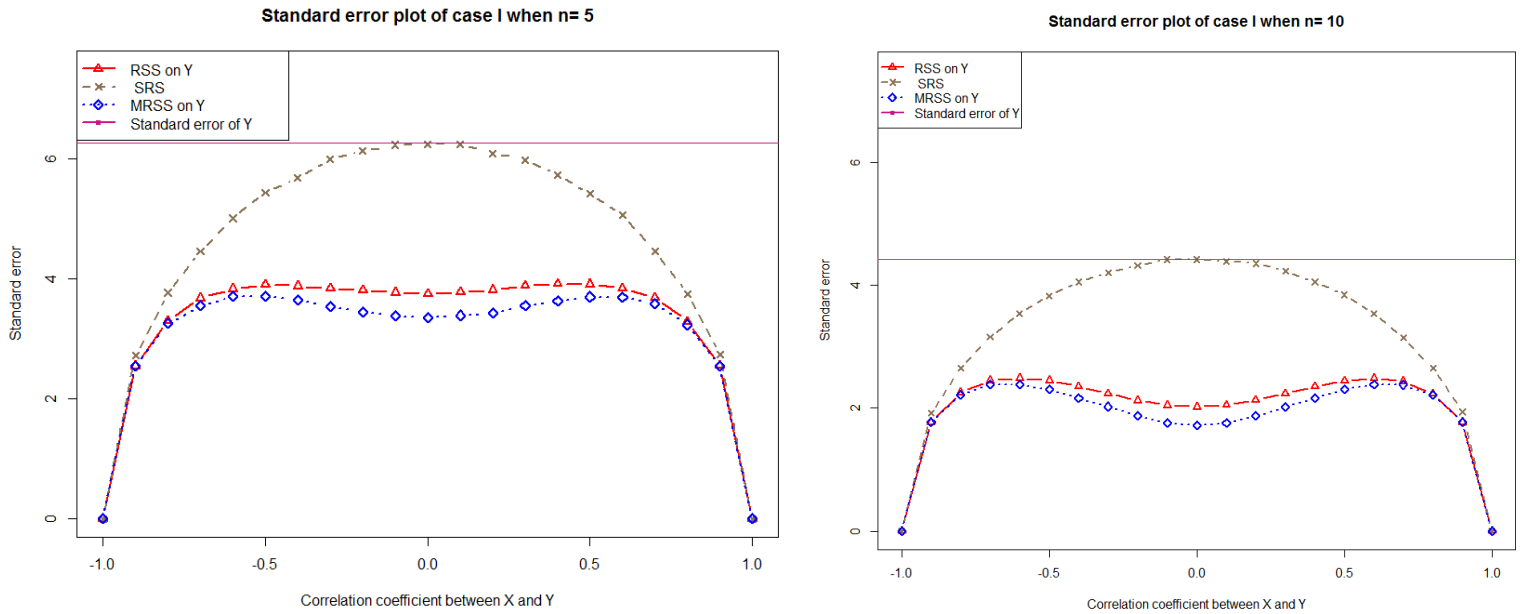


Figure 8.1: The estimated standard error plots of the regression estimator from the case I.

8.4 Simulation study

We assessed the performance of the proposed charts based on case I using ARL and standard deviation of the run length (SDRL). ARL_0 is the value of the ARL when a process is in-control (IC), while ARL_1 is the

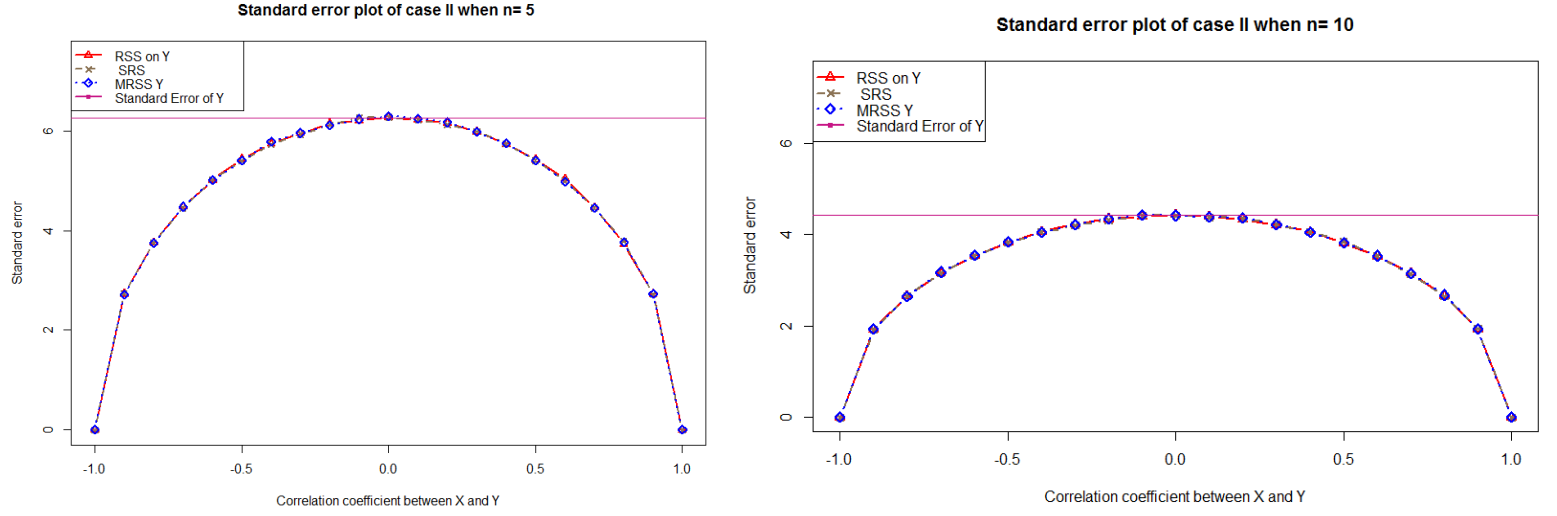


Figure 8.2: The estimated standard error plots of the regression estimator from the case II.

value of the ARL when the process is out-of-control (OOC) (Adegoke et al., 2018b). We considered small to moderate values of ϕ (i.e., $\phi \in \{0.05, 0.10, \text{or } 0.25\}$), and used corresponding values of that fixed ARL_0 to 500. This ensured a fair comparison of the proposed charts with other existing charts (i.e., the classical EWMA and the M_X EWMA control charts), using the same ARL_0 . The ARL values of the classical EWMA control chart with time-varying limits are given in Table 8.1. We define the size of shifts that we are interested in detecting, as $\delta = \frac{|\mu_1 - \mu_Y|}{(\sigma_Y/\sqrt{n})}$, where μ_1 is the OOC mean. The run length properties of the charts were investigated for different shifts, we considered $\delta \in \{0, 0.25, 0.5, 0.75, 1, 1.5, 2, 2.5, 3, 4, 5\}$.

Table 8.1: ARL values for the classical EWMA chart with $ARL_0 = 500$

ϕ	L	δ										
		0	0.25	0.5	0.75	1	1.5	2	2.5	3	4	5
0.05	2.639	500	77.75	23.71	11.87	7.31	3.77	2.43	1.77	1.41	1.09	1.01
0.10	2.824	500	103.3	28.81	13.61	8.21	4.17	2.66	1.92	1.51	1.12	1.01
0.25	3.001	500	169.5	47.38	19.32	10.41	4.78	2.94	2.09	1.62	1.16	1.20

The simulation study and methodology of the proposed charts are listed for both Phase I and Phase II as follows:

- In Phase I:

- For each sampling scheme, we generated n pair of bivariate normally distributed random vari-

able with parameters: $(\mu_Y = 100, \mu_X = 50, \sigma_Y = 14, \sigma_X = 5, \rho_{XY} : \rho_{XY} \in \{0.05, 0.25, 0.50, 0.75, \text{or } 0.95\})$.

n is the sample size at each time i . We used $n \in \{3, 5, 7, \text{or } 10\}$, this enabled us to examine the performance of the proposed charts with different sample sizes.

- ii We computed the regression estimators in Equations (8.6) and (8.10), from the n bivariate samples.
- iii We repeated step (i) and (ii) 20,000 times and computed the estimated parameters, i.e., the estimated mean and standard error of the regression estimators in Equations (8.7) and (8.11). These were used to set control limits in phase II.

- In Phase II:

- i At each time i , we generated a pair of a random variable (y_i, x_i) from a bivariate population with $(Y, X) \sim N_2(\mu_Y + (\sigma_Y / \sqrt{n}), \mu_X, \sigma_Y^2, \sigma_X^2, \rho_{XY})$.
- ii We computed the plotting statistics: W_i and P_i , given in Equations (8.8) and (8.12), respectively.
- iii Using the estimated parameters from Phase I, we constructed the estimated time-varying control limits using Equations (8.9) and (8.13), and compared W_i and P_i , against the set control limits.
- iv If W_i or P_i fell within the control limits, the process was considered to be IC, and the process repeated steps (i–iii) of Phase II, for the monitoring of the next test sample $i+1$. Alternatively, if W_i or P_i fell outside the control limits, the process was considered to be shifted to an OOC state. Consequently, the monitoring process was terminated for the chart, and we recorded the iteration number (or run length) that gave the first OOC signal.

We repeated Phase I and Phase II process 100,000 times. The average of the run lengths across simulations

(i.e., ARL) was reported. The ARL values are given in Tables 8.2 - 8.3, and the SDRL values are given in Tables 8.4 - 8.5, for $n \in \{3, 10\}$.

Table 8.2: ARL values of the proposed charts when $n = 3$.

Charts	ρ_{XY}	(ϕ, L)	δ										
			0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	4.00	5.00
S _X EWMA	0.05	(0.05, 2.639)	501.32	44.68	13.63	6.87	4.29	2.30	1.56	1.22	1.07	1.00	1.00
		(0.10, 2.824)	500.43	59.45	15.96	7.78	4.79	2.53	1.68	1.28	1.09	1.00	1.00
		(0.25, 3.001)	499.60	101.38	23.36	9.73	5.56	2.77	1.81	1.35	1.13	1.01	1.00
	0.25	(0.05, 2.639)	500.18	46.16	13.80	6.93	4.32	2.31	1.56	1.22	1.07	1.00	1.00
		(0.10, 2.824)	499.73	56.74	15.55	7.61	4.70	2.49	1.66	1.27	1.09	1.00	1.00
		(0.25, 3.001)	500.72	98.45	23.00	9.60	5.49	2.75	1.80	1.34	1.12	1.01	1.00
	0.50	(0.05, 2.639)	501.30	42.58	12.81	6.46	4.04	2.18	1.49	1.18	1.05	1.00	1.00
		(0.10, 2.824)	499.86	54.48	14.73	7.22	4.48	2.37	1.59	1.23	1.07	1.00	1.00
		(0.25, 3.001)	501.59	93.63	21.25	8.95	5.18	2.61	1.71	1.29	1.10	1.00	1.00
	0.75	(0.05, 2.639)	499.78	31.49	9.59	4.90	3.09	1.72	1.24	1.05	1.01	1.00	1.00
		(0.10, 2.824)	500.07	39.22	10.87	5.45	3.42	1.87	1.30	1.08	1.01	1.00	1.00
		(0.25, 3.001)	499.43	66.17	14.59	6.42	3.84	2.03	1.38	1.11	1.02	1.00	1.00
	0.95	(0.05, 2.639)	500.25	10.31	3.30	1.82	1.29	1.01	1.00	1.00	1.00	1.00	1.00
		(0.10, 2.824)	499.93	11.56	3.61	1.96	1.36	1.02	1.00	1.00	1.00	1.00	1.00
		(0.25, 3.001)	498.78	15.83	4.08	2.13	1.44	1.03	1.00	1.00	1.00	1.00	1.00
T _X EWMA	0.05	(0.05, 2.639)	499.99	38.94	11.86	6.01	3.77	2.06	1.42	1.14	1.03	1.00	1.00
		(0.10, 2.824)	501.45	50.92	13.80	6.82	4.23	2.25	1.52	1.19	1.05	1.00	1.00
		(0.25, 3.001)	500.72	85.60	19.59	8.26	4.83	2.46	1.63	1.24	1.07	1.00	1.00
	0.25	(0.05, 2.639)	503.38	39.31	12.03	6.12	3.82	2.08	1.43	1.14	1.04	1.00	1.00
		(0.10, 2.824)	498.24	50.09	13.77	6.82	4.23	2.26	1.53	1.19	1.05	1.00	1.00
		(0.25, 3.001)	500.60	88.86	20.08	8.44	4.89	2.50	1.65	1.25	1.07	1.00	1.00
	0.50	(0.05, 2.639)	501.32	38.91	11.82	5.97	3.74	2.04	1.41	1.13	1.03	1.00	1.00
		(0.10, 2.824)	499.76	49.96	13.56	6.69	4.16	2.22	1.50	1.18	1.04	1.00	1.00
		(0.25, 3.001)	500.69	83.94	19.05	8.10	4.72	2.42	1.61	1.23	1.06	1.00	1.00
	0.75	(0.05, 2.639)	499.03	30.20	9.25	4.73	2.99	1.68	1.21	1.04	1.00	1.00	1.00
		(0.10, 2.824)	500.76	38.52	10.59	5.29	3.32	1.82	1.28	1.07	1.01	1.00	1.00
		(0.25, 3.001)	499.53	61.87	13.79	6.12	3.68	1.96	1.34	1.09	1.01	1.00	1.00
	0.95	(0.05, 2.639)	501.05	10.11	3.25	1.80	1.28	1.01	1.00	1.00	1.00	1.00	1.00
		(0.10, 2.824)	501.74	11.49	3.59	1.96	1.35	1.02	1.00	1.00	1.00	1.00	1.00
		(0.25, 3.001)	500.72	15.66	4.05	2.12	1.43	1.03	1.00	1.00	1.00	1.00	1.00

8.5 Results and discussion

We summarize the major findings (given in Tables 8.2 - 8.5) from our proposed charts as follows:

- The use of the RSS and MRSS improve the performance of the EWMA chart for all values of ρ_{XY} considered in this study. These improvements are evident from the ARL1 values in Tables 8.2 - 8.3.

Table 8.3: ARL values of the proposed charts when $n = 10$.

Charts	ρ_{XY}	(ϕ, L)	δ										
			0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	4.00	5.00
S _X EWMA	0.05	(0.05, 2.639)	501.96	20.42	6.33	3.28	2.15	1.29	1.04	1.00	1.00	1.00	1.00
		(0.10, 2.824)	499.05	24.48	7.07	3.64	2.33	1.36	1.06	1.00	1.00	1.00	1.00
		(0.25, 3.001)	499.02	38.76	8.69	4.10	2.55	1.45	1.09	1.01	1.00	1.00	1.00
	0.25	(0.05, 2.639)	500.33	23.02	7.10	3.66	2.37	1.39	1.08	1.01	1.00	1.00	1.00
		(0.10, 2.824)	500.47	27.40	7.89	4.03	2.57	1.47	1.11	1.01	1.00	1.00	1.00
		(0.25, 3.001)	499.07	45.99	10.10	4.66	2.86	1.58	1.14	1.02	1.00	1.00	1.00
	0.50	(0.05, 2.639)	499.52	28.03	8.61	4.40	2.80	1.59	1.17	1.03	1.00	1.00	1.00
		(0.10, 2.824)	499.57	34.29	9.71	4.88	3.09	1.71	1.22	1.04	1.00	1.00	1.00
		(0.25, 3.001)	499.18	58.61	12.88	5.73	3.45	1.86	1.28	1.06	1.01	1.00	1.00
	0.75	(0.05, 2.639)	500.52	26.79	8.21	4.20	2.69	1.53	1.14	1.02	1.00	1.00	1.00
		(0.10, 2.824)	499.94	32.68	9.21	4.64	2.93	1.65	1.19	1.03	1.00	1.00	1.00
		(0.25, 3.001)	499.71	55.25	12.03	5.41	3.29	1.77	1.24	1.05	1.00	1.00	1.00
	0.95	(0.05, 2.639)	499.64	9.88	3.19	1.77	1.26	1.01	1.00	1.00	1.00	1.00	1.00
		(0.10, 2.824)	500.48	11.17	3.50	1.92	1.33	1.01	1.00	1.00	1.00	1.00	1.00
		(0.25, 3.001)	501.36	15.16	3.95	2.08	1.41	1.02	1.00	1.00	1.00	1.00	1.00
T _X EWMA	0.05	(0.05, 2.639)	499.21	15.46	4.85	2.57	1.72	1.12	1.01	1.00	1.00	1.00	1.00
		(0.10, 2.824)	499.70	18.14	5.42	2.82	1.86	1.16	1.01	1.00	1.00	1.00	1.00
		(0.25, 3.001)	499.79	27.32	6.38	3.12	2.01	1.21	1.02	1.00	1.00	1.00	1.00
	0.25	(0.05, 2.639)	500.27	18.86	5.82	3.05	2.00	1.23	1.03	1.00	1.00	1.00	1.00
		(0.10, 2.824)	501.49	22.82	6.60	3.39	2.19	1.30	1.04	1.00	1.00	1.00	1.00
		(0.25, 3.001)	500.76	34.84	7.90	3.77	2.37	1.37	1.06	1.00	1.00	1.00	1.00
	0.50	(0.05, 2.639)	500.08	24.84	7.72	4.00	2.57	1.48	1.12	1.02	1.00	1.00	1.00
		(0.10, 2.824)	500.42	31.61	8.91	4.47	2.84	1.59	1.16	1.02	1.00	1.00	1.00
		(0.25, 3.001)	499.42	51.88	11.35	5.15	3.14	1.72	1.21	1.04	1.00	1.00	1.00
	0.75	(0.05, 2.639)	500.63	26.00	7.94	4.06	2.61	1.49	1.12	1.02	1.00	1.00	1.00
		(0.10, 2.824)	500.16	31.20	8.86	4.47	2.85	1.60	1.17	1.03	1.00	1.00	1.00
		(0.25, 3.001)	500.04	52.75	11.50	5.21	3.17	1.73	1.22	1.04	1.00	1.00	1.00
	0.95	(0.05, 2.639)	500.33	9.93	3.18	1.77	1.26	1.01	1.00	1.00	1.00	1.00	1.00
		(0.10, 2.824)	501.82	11.27	3.51	1.92	1.33	1.01	1.00	1.00	1.00	1.00	1.00
		(0.25, 3.001)	500.21	15.39	3.97	2.08	1.41	1.02	1.00	1.00	1.00	1.00	1.00

- For small values of ρ_{XY} , the proposed T_XEWMA chart has smaller ARL1 values (and smaller SDRL values, cf. Tables 8.4 - 8.5) than the proposed S_XEWMA chart, especially for small values of δ . That is, when there is a small-to-moderate correlation between the variables, the T_XEWMA chart can detect small to moderate shifts quicker than the S_XEWMA chart. However, the performance of the charts is similar when a large value of ρ_{XY} is used (cf. Tables 8.2 - 8.3).
- The charts are ARL unbiased for any combinations of ϕ , L and ρ_{XY} . That is, the charts' ARL₀ values are bigger than their ARL₁ values, for any choice of δ (cf. Tables 8.2 - 8.3).

Table 8.4: SDRL values of the proposed charts when $n = 3$.

Charts	ρ_{XY}	(ϕ, L)	δ										
			0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	4.00	5.00
S _X EWMA	0.05	(0.05, 2.639)	498.04	37.15	9.54	4.42	2.56	1.20	0.71	0.44	0.25	0.05	0
		(0.10, 2.824)	499.86	53.51	11.47	4.89	2.78	1.29	0.76	0.49	0.30	0.06	0
		(0.25, 3.001)	499.98	98.33	19.97	7.01	3.45	1.42	0.82	0.54	0.34	0.08	0
	0.25	(0.05, 2.639)	500.00	38.54	9.65	4.45	2.59	1.21	0.70	0.44	0.25	0.04	0
		(0.10, 2.824)	497.68	50.34	11.15	4.78	2.73	1.28	0.75	0.48	0.29	0.06	0
		(0.25, 3.001)	498.43	95.47	19.60	6.82	3.38	1.42	0.82	0.53	0.33	0.07	0
	0.50	(0.05, 2.639)	499.68	35.09	8.88	4.12	2.39	1.12	0.66	0.40	0.21	0.03	0
		(0.10, 2.824)	498.43	48.21	10.42	4.48	2.58	1.19	0.71	0.45	0.25	0.04	0
		(0.25, 3.001)	500.00	90.74	17.92	6.29	3.14	1.32	0.77	0.50	0.30	0.06	0
	0.75	(0.05, 2.639)	497.99	24.85	6.42	3.01	1.74	0.82	0.46	0.23	0.09	0.01	0
		(0.10, 2.824)	500.00	33.13	7.27	3.23	1.87	0.88	0.51	0.27	0.11	0.01	0
		(0.25, 3.001)	499.10	62.94	11.43	4.13	2.15	0.96	0.56	0.31	0.13	0.01	0
	0.95	(0.05, 2.639)	498.74	6.97	1.88	0.88	0.50	0.11	0.01	0	0	0	0
		(0.10, 2.824)	499.73	7.81	2.00	0.94	0.55	0.13	0.01	0	0	0	0
		(0.25, 3.001)	498.32	12.69	2.34	1.02	0.60	0.16	0.01	0	0	0	0
T _X EWMA	0.05	(0.05, 2.639)	499.93	31.92	8.20	3.82	2.23	1.04	0.61	0.36	0.18	0.02	0
		(0.10, 2.824)	499.23	44.78	9.63	4.18	2.40	1.12	0.66	0.41	0.22	0.03	0
		(0.25, 3.001)	498.43	82.58	16.22	5.67	2.87	1.22	0.71	0.46	0.26	0.04	0
	0.25	(0.05, 2.639)	497.65	31.88	8.32	3.86	2.24	1.05	0.61	0.36	0.19	0.02	0
		(0.10, 2.824)	497.32	44.01	9.66	4.18	2.41	1.13	0.67	0.41	0.22	0.03	0
		(0.25, 3.001)	496.04	85.48	16.79	5.81	2.92	1.25	0.73	0.46	0.27	0.04	0
	0.50	(0.05, 2.639)	498.32	31.65	8.12	3.76	2.19	1.03	0.60	0.35	0.17	0.02	0
		(0.10, 2.824)	497.49	43.80	9.40	4.10	2.36	1.10	0.65	0.40	0.21	0.03	0
		(0.25, 3.001)	498.30	80.97	15.77	5.55	2.79	1.20	0.70	0.44	0.25	0.03	0
	0.75	(0.05, 2.639)	497.42	23.78	6.19	2.88	1.68	0.79	0.43	0.21	0.07	0	0
		(0.10, 2.824)	498.23	32.47	7.05	3.12	1.80	0.85	0.49	0.25	0.09	0.01	0
		(0.25, 3.001)	498.32	58.60	10.75	3.88	2.04	0.92	0.53	0.29	0.11	0.01	0
	0.95	(0.05, 2.639)	499.00	6.80	1.85	0.87	0.50	0.10	0	0	0	0	0
		(0.10, 2.824)	498.23	7.73	1.98	0.94	0.54	0.13	0.01	0	0	0	0
		(0.25, 3.001)	498.34	12.51	2.31	1.01	0.59	0.16	0.01	0	0	0	0

- As δ increases, the ARL and SDRL values approach 1 and 0, respectively (cf. Tables 8.2 - 8.5).

That is, for all values ρ_{XY} of the proposed charts detect large shifts promptly (cf. Tables 8.2 - 8.3).

- For fixed values of ρ_{XY} and δ , the proposed charts are more efficient using smaller values of ϕ (cf. Tables 8.2 - 8.3).
- For a fixed value of δ , the efficiency did not monotonically increase with ρ_{XY} . The performance of the charts tended to decrease for values of in the interval $0.25 \leq \rho_{XY} \leq 0.75$ especially for large values of n (cf. Tables 8.2 - 8.3). For case I, the patterns of dependency of the RSS and MRSS

Table 8.5: SDRL values of the proposed charts when $n = 10$.

Charts	ρ_{XY}	(ϕ, L)	δ										
			0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	4.00	5.00
S _X EWMA	0.05	(0.05, 2.639)	499.03	15.14	4.02	1.87	1.1	0.5	0.2	0.05	0.01	0	0
		(0.10, 2.824)	498.74	19.11	4.38	2.02	1.17	0.55	0.25	0.06	0.01	0	0
		(0.25, 3.001)	499	35.26	6.04	2.35	1.28	0.6	0.28	0.08	0.01	0	0
	0.25	(0.05, 2.639)	497.64	17.33	4.59	2.13	1.25	0.58	0.27	0.09	0.02	0	0
		(0.10, 2.824)	499.49	21.84	5	2.28	1.33	0.63	0.31	0.11	0.03	0	0
		(0.25, 3.001)	500	42.42	7.33	2.74	1.49	0.69	0.36	0.14	0.03	0	0
	0.5	(0.05, 2.639)	498.04	21.77	5.69	2.66	1.55	0.73	0.39	0.17	0.05	0	0
		(0.10, 2.824)	498.34	28.54	6.36	2.84	1.66	0.79	0.44	0.21	0.07	0	0
		(0.25, 3.001)	499.01	54.96	9.83	3.58	1.88	0.85	0.49	0.25	0.09	0	0
	0.75	(0.05, 2.639)	498.3	20.6	5.4	2.5	1.47	0.69	0.36	0.14	0.03	0	0
		(0.10, 2.824)	497.32	26.72	5.99	2.68	1.55	0.74	0.4	0.18	0.05	0	0
		(0.25, 3.001)	498.3	51.71	9.07	3.33	1.76	0.8	0.46	0.21	0.07	0	0
	0.95	(0.05, 2.639)	497.89	6.66	1.81	0.85	0.48	0.1	0	0	0	0	0
		(0.10, 2.824)	498.2	7.52	1.92	0.91	0.53	0.12	0.01	0	0	0	0
		(0.25, 3.001)	499.01	12.04	2.24	0.99	0.58	0.15	0.01	0	0	0	0
T _X EWMA	0.05	(0.05, 2.639)	499.23	11.01	2.95	1.39	0.82	0.33	0.08	0.01	0	0	0
		(0.10, 2.824)	498.75	13.36	3.23	1.49	0.88	0.38	0.11	0.01	0	0	0
		(0.25, 3.001)	498.23	24.01	4.1	1.66	0.95	0.42	0.13	0.02	0	0	0
	0.25	(0.05, 2.639)	499.23	13.79	3.66	1.72	1	0.45	0.16	0.03	0	0	0
		(0.10, 2.824)	498.04	17.58	4.06	1.85	1.08	0.5	0.2	0.04	0.01	0	0
		(0.25, 3.001)	499	31.32	5.38	2.11	1.17	0.55	0.23	0.06	0.01	0	0
	0.5	(0.05, 2.639)	498.34	18.95	5.07	2.37	1.39	0.65	0.33	0.12	0.03	0	0
		(0.10, 2.824)	499.23	25.88	5.74	2.57	1.49	0.71	0.38	0.15	0.04	0	0
		(0.25, 3.001)	499.03	48.34	8.45	3.13	1.67	0.77	0.42	0.19	0.05	0	0
	0.75	(0.05, 2.639)	499	19.91	5.22	2.42	1.41	0.66	0.34	0.13	0.03	0	0
		(0.10, 2.824)	499	25.31	5.71	2.56	1.5	0.71	0.38	0.16	0.04	0	0
		(0.25, 3.001)	496.34	49.13	8.6	3.16	1.69	0.78	0.43	0.19	0.06	0	0
	0.95	(0.05, 2.639)	496.54	6.67	1.8	0.85	0.48	0.09	0.01	0	0	0	0
		(0.10, 2.824)	497.3	7.6	1.93	0.91	0.53	0.12	0.01	0	0	0	0
		(0.25, 3.001)	498.34	12.21	2.23	0.99	0.58	0.15	0.01	0	0	0	0

regression estimators on ρ_{XY} are shown in Figure 8.1.

- The ARL performance of the proposed charts depends on the size of the sample size (n). Specifically, as n increased, the OOC ARL values decreased, especially for small to moderate shifts. For example, Figure 8.3 shows the ARL values of the proposed S_XEWMA and T_XEWMA charts for different values of n . The smoothing parameter used was $\phi = 0.05$, and $L = 2.639$, the correlation between the variables was $\rho_{XY} = 0.05$. As shown in the Figure 8.3, the ARL performance of the proposed charts increased as n increased.

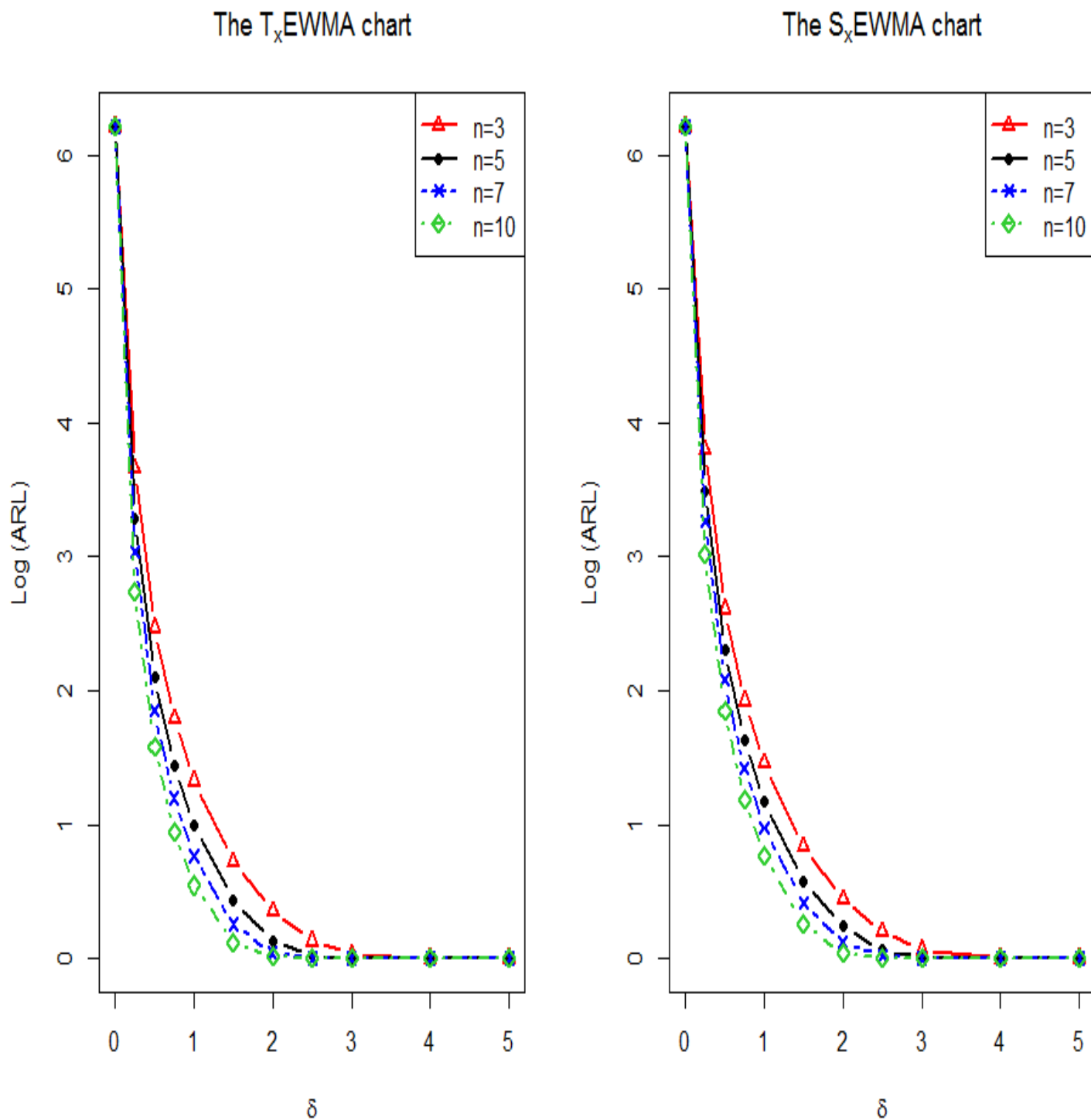


Figure 8.3: The plots of the proposed S_X EWMA and T_X EWMA charts for different values of n .

Abbas et al. (2014a) proposed the M_X EWMA control chart to improve the performance of the EWMA control chart when the process variable is observed along with an auxiliary variable. Tables 8.6 - 8.7 give

the ARL performance of the M_XEWMA chart for $n \in \{3, 10\}$, and corresponding values of ϕ and L given in Tables 8.1. The results in Tables 8.6 - 8.7, affirm that the M_XEWMA chart is more efficient in detecting shifts in the mean of a process than the classical EWMA chart (see Tables 8.1). The efficiency of the M_XEWMA chart over the classical EWMA is greater with larger values of ρ_{XY} . The performance of the M_XEWMA chart seems to be less dependent on sample size. Our proposed charts found that the application of RSS and MRSS improves the performance of the classical EWMA and M_XEWMA , especially when there is a small to a moderate correlation between process and auxiliary variables (i.e., $\rho_{XY} \leq 0.75$). This is shown by the lower ARL1 values of the SXEWMA and TXEWMA charts given in Tables 8.2 - 8.3.

Table 8.6: ARL values of the M_XEWMA chart for $n = 3$.

ρ_{XY}	(ϕ, L)	δ										
		0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	4.00	5.00
0.05	(0.05, 2.639)	499.96	77.48	23.71	11.84	7.38	3.76	2.42	1.77	1.41	1.09	1.01
	(0.10, 2.824)	501.44	103.32	29.36	13.74	8.30	4.19	2.66	1.92	1.51	1.12	1.01
	(0.25, 3.001)	501.66	168.28	47.80	19.46	10.47	4.77	2.95	2.09	1.62	1.16	1.02
0.25	(0.05, 2.639)	501.19	74.18	22.51	11.30	6.95	3.59	2.32	1.70	1.36	1.07	1.01
	(0.10, 2.824)	500.22	98.80	27.21	12.94	7.82	3.97	2.53	1.84	1.46	1.10	1.01
	(0.25, 3.001)	500.59	159.17	43.63	17.81	9.68	4.50	2.78	1.99	1.55	1.13	1.02
0.50	(0.05, 2.639)	501.96	61.23	18.52	9.33	5.76	3.01	1.97	1.48	1.22	1.02	1.00
	(0.10, 2.824)	500.29	80.47	21.93	10.58	6.44	3.32	2.16	1.59	1.28	1.04	1.00
	(0.25, 3.001)	501.96	138.50	34.99	14.24	7.86	3.74	2.35	1.71	1.36	1.05	1.00
0.705	(0.05, 2.639)	500.34	39.35	11.86	5.99	3.76	2.04	1.41	1.13	1.03	1.00	1.00
	(0.10, 2.824)	501.97	48.65	13.36	6.61	4.11	2.20	1.50	1.18	1.04	1.00	1.00
	(0.25, 3.001)	500.56	84.53	19.04	8.09	4.71	2.41	1.60	1.23	1.06	1.00	1.00
0.95	(0.05, 2.639)	500.09	10.70	3.42	1.88	1.32	1.02	1.00	1.00	1.00	1.00	1.00
	(0.10, 2.824)	500.50	12.17	3.77	2.04	1.40	1.02	1.00	1.00	1.00	1.00	1.00
	(0.25, 3.001)	501.28	16.87	4.27	2.22	1.48	1.04	1.00	1.00	1.00	1.00	1.00

The ARL performance measures the effectiveness of a chart at a particular shift point. We used the extra quadratic loss (EQL) and the relative average run length (RARL) to measure the overall performance of the charts. These measure the efficiency of the modified charts over all shifts (Abujiya, Mu'azu Ramat and Farouk, Abbas Umar and Lee, Muhammad Hisyam and Mohamad, 2013; Wu et al., 2009). The EQL

Table 8.7: ARL values of the M_X EWMA chart for $n = 10$.

ρ_{XY}	(ϕ, L)	δ										
		0.00	0.25	0.50	0.75	1.00	1.50	2.00	2.50	3.00	4.00	5.00
0.05	(0.05, 2.639)	500.69	77.42	23.62	11.85	7.30	3.76	2.42	1.77	1.41	1.09	1.01
	(0.10, 2.824)	501.38	103.00	829.39	13.68	8.27	4.20	2.66	1.92	1.51	1.12	1.01
	(0.25, 3.001)	499.66	167.60	46.82	19.18	10.32	4.75	2.92	2.08	1.61	1.16	1.02
0.25	(0.05, 2.639)	499.19	73.19	22.43	11.19	6.91	3.58	2.31	1.70	1.36	1.07	1.01
	(0.10, 2.824)	500.12	97.69	27.07	12.84	7.78	3.96	2.54	1.84	1.45	1.09	1.01
	(0.25, 3.001)	499.01	162.00	44.31	17.96	9.71	4.50	2.78	1.99	1.55	1.13	1.02
0.50	(0.05, 2.639)	499.59	61.19	18.53	9.29	5.74	3.00	1.97	1.47	1.21	1.02	1.00
	(0.10, 2.824)	500.43	81.72	22.16	10.63	6.47	3.33	2.15	1.59	1.28	1.04	1.00
	(0.25, 3.001)	500.56	133.90	34.40	14.09	7.75	3.71	2.35	1.70	1.35	1.05	1.00
0.75	(0.05, 2.639)	499.21	38.23	11.66	5.92	3.72	2.02	1.40	1.13	1.03	1.00	1.00
	(0.10, 2.824)	500.19	48.82	13.48	6.64	4.12	2.21	1.50	1.17	1.04	1.00	1.00
	(0.25, 3.001)	499.46	84.79	19.14	8.12	4.72	2.42	1.60	1.23	1.06	1.00	1.00
0.95	(0.05, 2.639)	499.58	10.56	3.40	1.87	1.31	1.02	1.00	1.00	1.00	1.00	1.00
	(0.10, 2.824)	499.91	12.02	3.74	2.03	1.39	1.02	1.00	1.00	1.00	1.00	1.00
	(0.25, 3.001)	500.96	16.85	4.26	2.22	1.49	1.04	1.00	1.00	1.00	1.00	1.00

is the weighted average of the ARL over all shifts and uses δ^2 as the weights. The EQL is given as:

$$EQL = \frac{1}{\delta_{max} - \delta_{min}} \int_{\delta_{min}}^{\delta_{max}} \delta^2 ARL(\delta) d(\delta). \quad (8.15)$$

where δ_{max} and δ_{min} are the maximum and the minimum values of shifts, respectively, $ARL(\delta)$ is the ARL of a particular chart at a shift size δ in the process mean. A chart with a smaller EQL is said to be better than its counterpart chart (Adegoke et al., 2017).

The RARL measures the overall performance of a chart on a benchmark chart and is calculated by:

$$RARL = \frac{1}{\delta_{max} - \delta_{min}} \int_{\delta_{min}}^{\delta_{max}} \frac{ARL(\delta)}{ARL_{benchmark}(\delta)} d(\delta). \quad (8.16)$$

$ARL_{benchmark}(\delta)$ is the ARL of the benchmark chart with shift δ . Charts with an RARL less than one are considered superior to the benchmark chart (Abbasi et al., 2015). The EQL and RARL were solved using numerical integration. The EQL and RARL of the charts with $\phi = 0.05$ and $L = 2.639$, for $n \in \{3, 5, 10\}$, are given in Tables 8.8 - 8.9, respectively.

Table 8.8: EQL values of the control charts.

ρ_{XY}	$n = 3$					$n = 5$					$n = 10$				
	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
Classical EWMA	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688	9.688
M_X EWMA	9.675	9.354	8.317	6.742	5.346	9.68	9.292	8.308	6.706	5.345	9.673	9.326	8.299	6.71	5.344
S_X EWMA	7.132	7.157	6.954	6.29	5.333	6.365	6.433	6.513	6.167	5.325	5.729	5.856	6.108	6.034	5.322
T_X EWMA	6.756	6.788	6.733	6.22	5.328	6.038	6.128	6.313	6.092	5.327	5.522	5.66	5.966	5.99	5.322

The EQL and RARL values in Tables 8 and 9 show that the classical EWMA and M_X EWMA control charts are inferior to the proposed charts for all values of ρ_{XY} , especially when $\rho_{XY} \leq 0.75$. Moreover, the efficacy of our new charting schemes increases with n .

Table 8.9: RARL values of the control charts.

ρ_{XY}	$n = 3$					$n = 5$					$n = 10$				
	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95	0.05	0.25	0.5	0.75	0.95
EWMA	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
M_X EWMA	1	0.961	0.836	0.636	0.436	0.999	0.952	0.835	0.63	0.435	0.998	0.957	0.834	0.631	0.435
S_X EWMA	0.687	0.691	0.664	0.575	0.434	0.585	0.594	0.605	0.557	0.432	0.495	0.513	0.549	0.539	0.432
T_X EWMA	0.637	0.641	0.634	0.565	0.433	0.539	0.552	0.577	0.547	0.432	0.463	0.484	0.529	0.533	0.432

8.6 Illustrative examples

We provide illustrative examples of the proposed charts using a simulated benchmark process: the continuous stirred tank reactor (CSTR) datasets. The non-isothermal CSTR was originally described in Marlin (2000) and has been widely used as a benchmark in fault detection and diagnosis (see Shi et al. (2013); Yoon and Macgregor (2001)). The CSTR process comprises of nine study variables, amongst which we chose the outlet temperature as the process variable (Y), and the inlet temperature as the auxiliary information (X).

The dataset originally contained 1024 samples; the first 512 values occurred when the process was in an IC state. These are used as the Phase I dataset. The parameters of the Phase I sample are: $\mu_Y = 368.2328, \sigma_Y^2 = 0.2185915, \mu_X = 369.8789, \sigma_X^2 = 0.3180327$ and the correlation between the variables

is $\rho_X Y = 0.08974039$. These parameters were used to estimate the regression estimator from a sample of size $n=5$, and the control charts limits parameters were estimated. Considering these estimates as known parameters, we generated 20 paired observations of size $n = 5$ from a bivariate normal distribution using the different sampling schemes considered in this study. A shift (δ) was introduced to the mean of the process variable after the fifth observations. This shifted the process mean to $\mu_1 = 368.4419$.

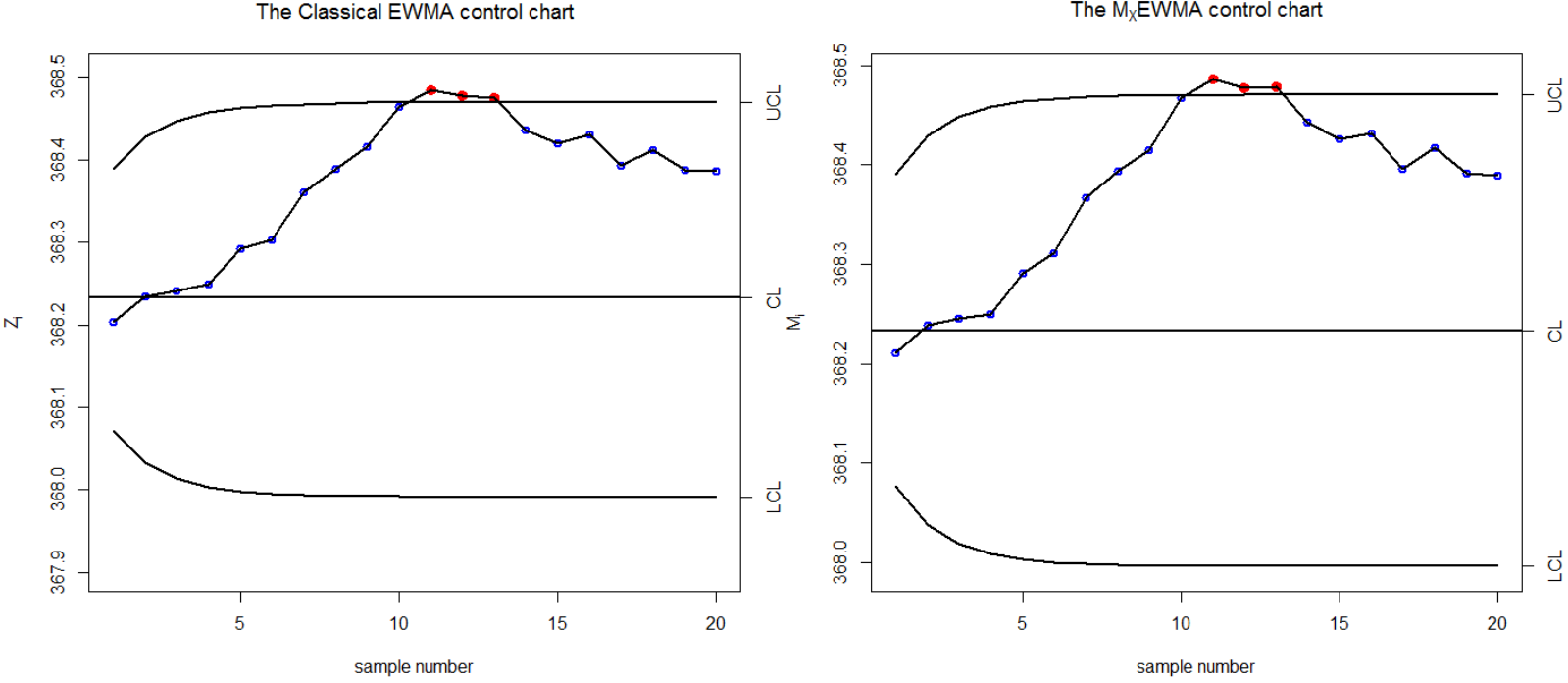


Figure 8.4: The Classical EWMA and the M_X EWMA control charts CSTR dataset.

The classical EWMA and M_X EWMA charts detect the first OOC signal after the 10th sample (Figure 8.4). However, the proposed charts: S_X EWMA and T_X EWMA, give an initial OOC signal after the 7th sample (see Figure 8.5).

8.7 Summary and conclusion

In this chapter, we proposed improvements to the performance of the classical EWMA control chart for those scenarios where measuring the process variable is expensive, but ranking of units according to process

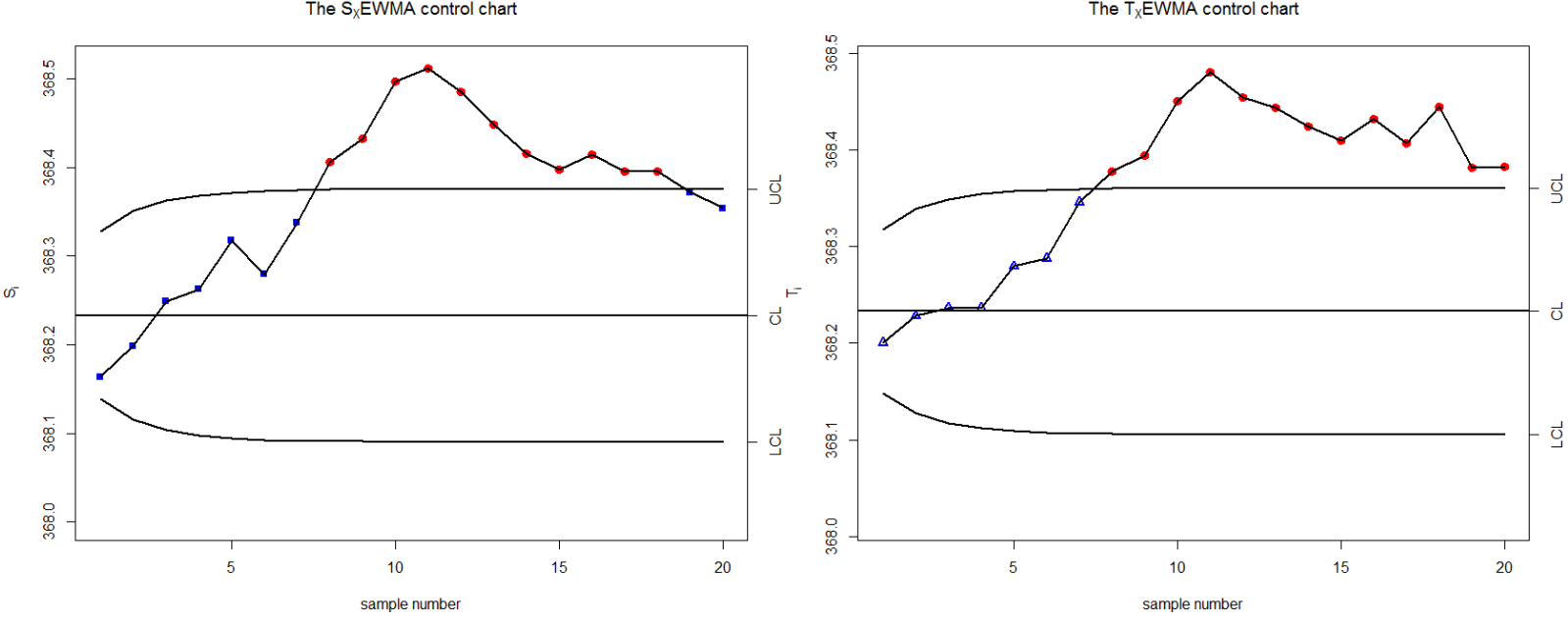


Figure 8.5: The S_X EWMA and T_X EWMA control charts for the CSTR dataset.

variable is relatively easy and cheap, and the process variable is observed with auxiliary information. We developed control charts based on RSS, and MRSS and compared the results of our proposed charts: the S_X EWMA chart and the T_X EWMA charts, with the results of the charts based on SRS (i.e., M_X EWMA) and the classical EWMA. The performance of the proposed chart was evaluated in terms of ARL and overall performance measures. We showed, through simulation, that our proposed charts are more efficient than the classical EWMA and the M_X EWMA charts in detecting a shift in the location parameter of the process variable, even when there is only a small to moderate correlation between the variables. Our simulation results also showed that the proposed T_X EWMA charts are more efficient than the proposed S_X EWMA chart, especially for small to moderate shifts in the location parameter.

Chapter 9

Efficient monitoring of a process mean using an auxiliary variable under simple random sampling

”This is the peer reviewed version of the following article: Adegoke, N. A., Smith, A. N., Anderson, M. J., Sanusi, R. A., & Pawley, M. D. (2019). Efficient Homogeneously Weighted Moving Average Chart for Monitoring Process Mean Using an Auxiliary Variable. IEEE Access, 7, 94021-94032, which has been published in final form at <https://doi.org/10.1109/ACCESS.2019.2926533>. This article may be used for non-commercial purposes in accordance with IEEE Terms and Conditions for Self-Archiving.”

In this chapter, we propose an efficient control chart for monitoring small shifts in a process mean for scenarios where the process variable is observed with a correlated auxiliary variable. The proposed chart, called an AHWMA chart, is an homogeneously weighted moving average type control chart that uses both the process and auxiliary variables in the form of a regression estimator to provide an efficient and

unbiased estimate of the mean of the process variable. We provide the design structure of the chart and examine its performance in terms of its run length properties. Using a simulation study, we compare its run length performance with several existing methods for detecting a small shift in the process mean. Our simulation results show that the proposed chart is more efficient in detecting a small shift in the process mean than its competitors. We provide a detailed study of the chart's robustness to non-normal distributions and shown that the chart may also be designed to be less sensitive to non-normality. We give some recommendations on the application of the chart when the process parameters are unknown, and provide an example to show the implementation of the proposed new technique.

9.1 Introduction

Monitoring programmes are designed to detect unnatural changes in process variables for a wide variety of applications, particularly in industrial and manufacturing settings. Control charts are the most popular and sophisticated tools for tracking processes of interest, ensuring they are kept in control by monitoring essential quality characteristics (Yen and Shiau, 2010). To date, several univariate control charts have been proposed in statistical process control (SPC) literature; they are classified into (i) memory-less control charts and (ii) memory-type control charts; these are useful for monitoring large and small-to-moderate shifts in the process, respectively. For example, the Shewhart chart is a memory-less type control chart that uses only the current process information without referring to past behavior of the process. It is very effective for detecting a large shift in the process mean (i.e., $\delta \geq 2$, where δ is the size of the shift in standard deviation units (Testik et al., 2003)). The homogeneously weighted moving average (HWMA) control chart by Abbas (2018), is a memory-type chart proposed for efficient monitoring of small (i.e., $\delta \leq 0.5$) to moderate (i.e., $0.5 < \delta < 2$) shifts in the process mean. Other memory-type charts include the EWMA chart by Roberts (1959), the CUSUM chart by Page (1961), and the mixed EWMA-CUSUM

chart proposed by Abbas et al. (2013).

These univariate classical charts are widely used in most of today's industries; their attractiveness is motivated by their simplicity of construction, implementation, and interpretation, as well as their prompt detection of small, moderate, or large shifts in a process mean. These techniques have been implemented by Abtew et al. (2018) to monitor the quality of garments produced on the sewing floor, by El-Din et al. (2006) to monitor and control steam boiler generation for vacuum degassing processes, and by Hayes et al. (1997) to evaluate critical control point hygiene data. Also, see Benneyan (1998a,b); Srikaeo and Hourigan (2002); De Vries and Conlin (2003), and Madsen and Kristensen (2005) for some other industrial applications of these classical charts.

Several applications of classical charts focus on monitoring the process in situations where the process variable is independent of other variables; however, in some cases, the process variable may be observed along with another correlated auxiliary variable. The concept of using supplemental information to provide an efficient estimate of a population parameter is popular in the field of survey sampling (Cochran, 1977). Several researchers have studied and recommended the introduction and application of auxiliary variables into the monitoring scheme of a process variable of interest, and have proposed a variety of different control charts tools for this purpose.

For example, Mandel (1969) proposed a regression control chart, while Zhang (1985) proposed a cause-selecting control chart. Recently, Riaz (2008b) proposed a Shewhart-type chart in the form of a regression-based estimator, called a V_r chart, for monitoring process variability. He compared the proposed V_r chart with some other existing charts (specifically, R , S and S^2 charts for the same purpose), and showed that the V_r chart was effective in detecting moderate to large shifts in the process variability under certain conditions on the correlation between the process variable and auxiliary variable. Similarly, a Shewhart-type control chart using a regression-based estimator (M_r chart) for monitoring a process mean (proposed by Riaz (2008a)), was shown to be more powerful at detecting shifts in the process mean.

This work was later extended to an EWMA chart for detecting small-to-moderate changes in the process mean under different correlation structures between the process and auxiliary variables (see Adegoke et al. (2018a, 2017); Ridwan A. Sanusi (2017); Abbas et al. (2014b)).

Here, we propose a more efficient control chart for monitoring the process mean when the process variable is observed along with an auxiliary variable. The proposed chart, called an auxiliary homogeneously weighted moving average (AHWMA) chart, is an HWMA-type control chart that uses both the deviation of the process mean from its target value (whether known apriori or estimated from historical reference samples), as well as a regression estimator for the process mean provided through its relationship (or estimated relationship) with an auxiliary variables with which it is known to be correlated. The rest of the chapter is organized as follows: in Section 9.2, we provide the design structure of the chart. Section 9.3 compares the AHWMA chart (run length) performance in detecting a small shift in the process mean with several other existing charts. Section 9.4 gives a detailed study of the chart's robustness to non-normality. We give recommendations regarding the application of the chart when the process parameters are unknown in Section 9.5. Section 9.6 provides an example to demonstrate practical implementation of the chart, followed by a conclusion and discussion in Section 9.7.

9.2 The AHWMA Control Chart

Consider constructing control chart based on observations z_{ij} of the quality characteristics Z_{ij} , for each of $i = 1, \dots, m$ time-points and $j = 1, \dots, n$ sampling units per time-point (i.e., n is the sample size). Assume that these quality characteristics (Z_{ij}), are identically distributed as normal random variables with an in-control known mean (μ_Z) and standard deviation (σ_Z), i.e., $Z_{ij} \sim N(\mu_Z, \sigma_Z^2)$ and represents the main process variable. The HWMA statistic, H_i (in Equation (9.1)), at time-point i , gives a specific weight to the current sample and the remaining weight is equally distributed among the previous samples,

and is given as:

$$H_i = w\bar{z}_i + (1 - w)\bar{\bar{z}}_{i-1} \quad (9.1)$$

where \bar{z}_i is the sample average for the i th sample, and w is a smoothing constant (also called the sensitivity parameter) selected such that $0 < w \leq 1$. The HWMA structure becomes the Shewhart plotting structure whenever $w = 1$. $\bar{\bar{z}}_{i-1}$ is the average of the sample means of all of the previous samples (i.e., up to and including the $(i - 1)$ th sample), and is given as $\bar{\bar{z}}_{i-1} = \frac{1}{n} \sum_{k=1}^{i-1} \bar{z}_k$. The mean and variance of the HWMA statistic in Equation (9.1) are given as $\mu_H = \mu_Z$, and

$$\sigma_{H_i}^2 = \begin{cases} \frac{1}{n} w^2 \sigma_Z^2 & \text{if } i = 1 \\ \frac{1}{n} \left(w^2 \sigma_Z^2 + (1 - w)^2 \frac{\sigma_Z^2}{i - 1} \right) & \text{if } i > 1 \end{cases} \quad (9.2)$$

where $\mu_H = \mu_Z$ and σ_Z^2 are the mean and variance of the normally distributed random variable Z (Abbas, 2018).

Let an auxiliary variable Y_{ij} be correlated with the main variable of interest, Z_{ij} , with correlation ρ . We assume the observations of Z_{ij} and Y_{ij} are observed in pairs from a bivariate normal distribution, given as, $(Z, Y) \sim N_2(\mu_Z, \mu_Y, \sigma_Z^2, \sigma_Y^2, \rho)$, where N_2 is the bivariate normal distribution, μ_Y and σ_Y^2 , are the the population mean and variance of Y , respectively. We assume the linear relationship between the variables can be modelled using a linear least squares obtained by adjusting the process mean at time i , z_i , to reflect its known relationship with the auxiliary variable, yielding the regression-informed estimator (i.e., R_i) for the process mean given as:

$$R_i = \bar{z}_i + b(\mu_Y - \bar{y}_i) \quad (9.3)$$

where b is the slope of the regression line; given as the change in the process variable, Z , due to a unit

change in the auxiliary variable, Y , (Cochran, 1977). The mean and variance of R are given as:

$$\mu_R = \mu_Z \text{ and } \sigma_R^2 = \frac{\sigma_Z^2}{n}(1 - \rho^2), \quad (9.4)$$

respectively.

Using Equation (9.3), the plotting statistic (T_i) of the AHWMA chart is given as:

$$T_i = wR_i + (1 - w)\bar{R}_{i-1} \quad (9.5)$$

where w is the smoothing parameter of the chart (selected such that $0 \leq w \leq 1$), R_i is the regression-informed estimate of the process variable, given in Equation (9.3) for the i th sample, and \bar{R}_{i-1} is the average of the sample means of all of the previous samples (i.e., up to and including the $(i - 1)$ th sample) of the plotting statistic, and is given as $\bar{R}_{i-1} = \frac{1}{n} \sum_{k=1}^{i-1} R_k$. The mean and variance of the plotting statistic in Equation (9.5) are given as $\mu_H = \mu_Z$ (also called the centre line of the AHWMA chart), and

$$\sigma_{T_i}^2 = \begin{cases} \frac{(1 - \rho^2)}{n} w^2 \sigma_Z^2 & \text{if } i = 1 \\ \frac{(1 - \rho^2)}{n} \left(w^2 \sigma_Z^2 + (1 - w)^2 \frac{\sigma_Z^2}{i - 1} \right) & \text{if } i > 1, \end{cases} \quad (9.6)$$

respectively. The time varying lower (L_i) and upper (U_i) control chart limits of the plotting statistic given in Equation (9.5) are given as:

$$L_i = \begin{cases} \mu_Z - C\sigma_Z \sqrt{\frac{w^2}{n}(1 - \rho^2)} & \text{if } i = 1 \\ \mu_Z - C\sigma_Z \sqrt{\left(\frac{w^2}{n} + \frac{(1 - w)^2}{n(i - 1)} \right) (1 - \rho^2)} & \text{if } i > 1 \end{cases} \quad (9.7)$$

and

$$U_i = \begin{cases} \mu_Z + C\sigma_Z \sqrt{\frac{w^2}{n}(1 - \rho^2)} & \text{if } i = 1 \\ \mu_Z + C\sigma_Z \sqrt{\left(\frac{w^2}{n} + \frac{(1 - w)^2}{n(i - 1)}\right)(1 - \rho^2)} & \text{if } i > 1, \end{cases} \quad (9.8)$$

respectively, where, C determines the width of the control limits; the values of C and w are chosen to achieve a desired in-control ARL for the chart. We provide R-code (R Core Team, 2013) (in the appendix C.3.1) which practitioners can use to obtain the value of C , given w , that fix the in-control ARL of the chart to a desired value. We adopted the ARL numerics algorithm for the EWMA chart Knoth (2017); implemented in the `spc (R)` package Knoth (2018), to obtain an arbitrary start value (say C^{start}) of the AHWMA chart limit, and used a binary search algorithm to search for the corrected limit (C) for the chart.

9.3 Performance assessments and comparisons

Performance assessments

Here, we provide a comprehensive assessment of the AHWMA chart in detecting a shift in the process mean in terms of the chart's average run length (ARL) and standard deviation of run length (SDRL). ARL is the average number of plotted samples on the control chart before a shift is detected. The in-control ARL, denoted by ARL_0 , is the value of the ARL when a process is in control, while the out-of-control ARL, denoted by ARL_1 , is the value of the ARL when the process is out of control. $SDRL$ is used to determine the variation of the run length distribution for a given value of shift. Similarly, $SDRL_0$ and $SDRL_1$ can be defined as $SDRL$ for the in-control and out-of-control process, respectively. When comparing two charts, the ARL_0 is fixed to a specific value, and a chart having a smaller value of ARL_1 than another is said to be more efficient in detecting the shift in the process (Adegoke et al., 2019; Abbasi

et al., 2018; Abbas et al., 2018; Khan et al., 2018).

To ensure a fair comparison of the AHWMA chart with existing charts of the same ARL_0 , we examined the performance of the chart with $w \in \{0.03, 0.05, 0.10, 0.25, 0.5, 0.75\}$, and the corresponding values of C that fix ARL_0 to 500 are used, the R-code provided in the appendix C.3.1 finds the value of C (for each value of w), that fixes ARL_0 to 500. We examined the ARL performance of the chart under different correlation values between the process and the auxiliary variables; in particular, we considered $\rho \in \{0.05, 0.25, 0.5, 0.75, 0.95\}$. The ARL values of the AHWMA chart are given in Tables 9.1 - 9.5. In these tables, δ is the size of shifts, and is calculated $\delta = \frac{n^{1/2}|\mu_Z - \mu_1|}{\sigma_Z}$, where n is the sample size at each time i (here, we assume $n = 1$ across i), and μ_Z and μ_1 are the in-control and out-of-control mean, respectively.

Table 9.1: ARL and $SDRL$ values of the AHWMA chart when the correlation between the variables is $\rho = 0.05$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w .

	w											
	0.03		0.05		0.1		0.25		0.5		0.75	
δ	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
0.000	502.98	428.85	498.7	371.79	502.95	410.23	501.41	483.52	497.28	495.4	500.09	495.64
0.050	363.45	326.86	382.76	296.57	396.35	325.16	438.68	420.99	478.99	475.17	488.54	490.08
0.075	273.65	250.5	295.14	231.71	317.18	255.34	384.78	368.41	452.28	449.05	476.29	473.72
0.100	206.62	187.81	228.92	180.9	250.84	199.77	326.26	309.14	417.58	417.27	454.29	452.94
0.125	159.27	141.92	180.75	140.23	198.63	154.81	272.98	256.67	383.76	385.05	436.42	435.91
0.150	126.43	111.15	144.81	110.84	162.37	123.83	226.05	209.25	346.85	345.27	411.53	413.32
0.175	102.59	89.67	119.59	89.15	132.35	98.6	185.95	169.49	309.12	304.8	389.93	388.13
0.200	84.61	72.93	100.41	74.29	111.43	80.63	156.47	139.98	276.56	274.03	362.66	359.46
0.250	60.59	51.6	72.93	52.75	81.49	56.88	112.27	97.75	217.59	213.85	313.11	312.89
0.500	20.05	15.71	24.89	17.02	28.54	17.71	33.81	25.21	68.28	64.33	132.08	131.2
0.750	10.31	7.41	12.74	8.25	14.88	8.74	16.13	10.66	27.82	24.51	57.82	56.92
1.000	6.57	4.25	8.01	4.82	9.33	5.16	9.67	5.8	14.09	11.49	28.26	26.74
1.500	3.74	2.13	4.42	2.31	4.96	2.43	4.93	2.54	5.62	3.82	9.19	8.01
2.000	2.55	1.45	2.98	1.52	3.31	1.52	3.19	1.48	3.2	1.82	4.21	3.18
	C											
	2.272		2.608		2.938		3.075		3.089		3.09	

The main findings of the AHWMA chart (cf. Tables 9.1 - 9.5) are:

- For fixed values of δ and ρ , the chart is more efficient for smaller value of w . For example, where

$\rho = 0.05$ (Table 9.1), when $\delta = 0.5$, the values of the ARL_1 when $w = 0.03$ and 0.75 were 20.05 and

Table 9.2: ARL and $SDRL$ values of the AHWMA chart when the correlation between the variables is $\rho = 0.25$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w .

	w											
	0.03		0.05		0.1		0.25		0.5		0.75	
δ	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
0	502.5	429.42	498.92	372.53	502.19	411.15	501.01	488.82	499.07	492.39	500.98	496.82
0.050	358.21	321.07	378.19	292.9	391.86	321.41	437.22	418.88	476.3	469.57	489.51	490.58
0.075	264.91	242.15	288.01	226.75	311.24	252.59	376.49	360.73	451.69	446.04	471.6	472.85
0.100	199.08	180.8	221.19	173.23	243.62	193.54	317.17	302.38	413.08	413.82	457.88	455.84
0.125	153.2	136.33	174.67	135.14	191.55	148.78	263.92	246.52	379.53	378.61	432.72	430.93
0.150	121.43	106.87	139.66	106.8	155.22	117.75	218.69	201.72	338.72	336.89	409.57	409.29
0.175	98.06	85.24	114.52	85.74	128.19	94.54	179.98	164.52	304.51	300.64	381.68	379.18
0.200	81.18	69.62	95.62	70.28	106.39	77.07	148.91	133.53	268.09	263.87	356.45	358.04
0.250	58.11	49.05	69.54	50.26	77.88	54.35	106.9	92.35	208.71	206.16	305	303.75
0.500	19.02	15	23.88	16.33	27.25	16.85	31.91	23.63	64.32	60.14	125.44	123.43
0.750	9.74	6.92	12.21	7.83	14.16	8.25	15.32	9.99	25.75	22.63	53.68	52.13
1.000	6.29	4	7.65	4.53	8.85	4.84	9.16	5.39	13	10.42	26.05	24.65
1.500	3.57	2.02	4.19	2.17	4.72	2.3	4.68	2.37	5.27	3.49	8.38	7.23
2.000	2.45	1.4	2.86	1.46	3.18	1.47	3.03	1.4	3.03	1.69	3.93	2.91
	C											
	2.272		2.608		2.938		3.075		3.089		3.09	

Table 9.3: ARL and $SDRL$ values of the AHWMA chart when the correlation between the variables is $\rho = 0.5$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w .

	w											
	0.03		0.05		0.1		0.25		0.5		0.75	
δ	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
0	502.09	428.04	498.36	371.62	501.47	406.7	504.88	486.43	498.55	497.01	496.55	498.91
0.050	333.66	301.2	354.43	276.58	373.92	302.68	425.24	405.64	471.88	466.32	487.38	486.56
0.075	240.95	220.02	263.31	207.91	285.65	231.4	354.88	339.89	435.63	429.98	469.55	466.55
0.100	174.83	156.92	197.66	153.49	216.45	170.81	290.14	273.78	397.54	393.99	445.34	446.92
0.125	132.93	117.26	151.91	115.44	168.43	129.25	234.95	218.49	357.17	354.94	419.21	419.54
0.150	104.06	90.67	120.7	91.47	135.03	100.63	190.51	174.88	312.66	310.1	390.07	387.53
0.175	83.71	71.86	98.45	72.9	110.08	80.03	153.84	138.3	275.08	270.88	363.64	363.96
0.200	69.32	59.15	81.65	59.65	91.18	65.05	127.81	112.89	239.37	235.41	331.84	333.34
0.250	49.08	41.15	59.23	42.45	65.89	44.93	89.86	76.3	180.91	176.47	276	276.42
0.500	15.75	12.08	19.73	13.38	22.63	13.72	25.77	18.3	50.43	46.59	101.74	100.72
0.750	8.23	5.64	10.12	6.38	11.79	6.7	12.52	7.87	19.87	16.95	40.79	39.3
1.000	5.35	3.28	6.44	3.67	7.43	3.97	7.59	4.31	10.06	7.75	19.11	17.68
1.500	3.1	1.73	3.61	1.84	4.04	1.89	3.95	1.91	4.21	2.63	6.12	4.98
2.000	2.1	1.23	2.43	1.3	2.72	1.29	2.61	1.19	2.5	1.29	3	2.03
	C											
	2.272		2.608		2.938		3.075		3.089		3.09	

132.08, respectively. Thus, the chart detects a shift in the process mean faster when a small value of w is used.

Table 9.4: ARL and $SDRL$ values of the AHWMA chart when the correlation between the variables is $\rho = 0.75$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w .

	w											
	0.03		0.05		0.1		0.25		0.5		0.75	
δ	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
0	501.65	430.15	498.97	374.39	499.57	404.52	504.97	490.17	493.17	493.87	497.6	500.41
0.050	271.72	248.22	297.52	234.08	315.26	256.33	385.37	369.96	450.49	449.65	476.24	477.56
0.075	178.17	160.77	201.34	156.45	222	174.13	296.36	278.87	398.8	400.65	450.12	449.58
0.100	125.55	110.22	144.11	109.47	160.49	122.96	223.57	207	343.09	341.7	410.87	406.97
0.125	91.66	79.47	108.24	79.75	120.11	88.2	171.01	155.54	290.88	287.62	375.59	373.21
0.150	70.54	60.2	84	61.41	93.52	66.57	131.78	115.65	245.44	239.44	336.11	332.07
0.175	55.94	47.27	67.61	49.08	75.41	52.08	103.9	89.92	201.67	198.62	298.92	297.8
0.200	45.3	37.77	54.88	39.25	61.84	41.94	83.18	70.42	169.04	164.22	262.48	262.96
0.250	31.95	25.93	39.23	27.61	44.38	28.8	55.98	44.86	117.61	113.89	203.97	203.6
0.500	10.24	7.26	12.66	8.17	14.71	8.57	15.96	10.46	27.41	24.28	56.59	55.52
0.750	5.5	3.42	6.68	3.83	7.64	4.08	7.79	4.42	10.53	8.18	19.97	18.55
1.000	3.69	2.11	4.35	2.28	4.91	2.39	4.86	2.49	5.53	3.72	8.94	7.78
1.500	2.16	1.26	2.52	1.33	2.8	1.31	2.67	1.21	2.57	1.35	3.11	2.14
2.000	1.45	0.85	1.65	0.96	1.86	0.99	1.76	0.84	1.66	0.73	1.76	0.92
	C											
	2.272		2.608		2.938		3.075		3.089		3.09	

Table 9.5: ARL and $SDRL$ values of the AHWMA chart when the correlation between the variables is $\rho = 0.95$. The values of C are chosen to fix the chart's ARL_0 to 500 for each chosen value of w .

	w											
	0.03		0.05		0.1		0.25		0.5		0.75	
δ	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$	ARL	$SDRL$
0	502.44	429.19	499.63	372.3	501.08	409.78	501.37	483.01	504.2	502.97	503.71	503.15
0.050	115.73	101.52	133.96	101.16	148.56	112.37	209.06	193.55	332.14	328.99	401.56	398.62
0.075	65.13	55.07	77.2	55.89	86.73	61.75	120.3	105.64	228.82	223.07	322.7	323.39
0.100	41.61	34.64	50.65	35.8	56.68	37.78	74.83	62.18	155.76	152.02	247.89	248.76
0.125	28.84	23.43	35.68	25.06	40.43	26.11	50.99	40.33	105.88	102	187.84	185.45
0.150	21.43	17.04	26.85	18.37	30.69	19.2	36.25	27.29	75.21	71.06	140.91	139.68
0.175	16.54	12.67	20.62	13.88	23.92	14.67	27.41	19.86	54.38	50.75	107.39	106.69
0.200	13.35	10.06	16.58	11.01	19.31	11.63	21.51	14.81	39.96	36.15	81.86	80.94
0.250	9.32	6.55	11.5	7.33	13.41	7.75	14.43	9.32	23.77	20.57	49.83	48.49
0.500	3.43	1.91	4.01	2.07	4.5	2.16	4.45	2.23	4.93	3.23	7.68	6.53
0.750	1.99	1.18	2.33	1.25	2.59	1.24	2.46	1.12	2.34	1.19	2.76	1.8
1.000	1.33	0.75	1.52	0.88	1.69	0.92	1.62	0.77	1.53	0.66	1.6	0.79
1.500	1.01	0.12	1.02	0.19	1.04	0.24	1.04	0.21	1.04	0.2	1.04	0.21
2.000	1	0	1	0.01	1	0.02	1	0.02	1	0.02	1	0.02
	C											
	2.272		2.608		2.938		3.075		3.089		3.09	

- For fixed values of δ , w and C , the chart is more efficient for large values of ρ are used. For example,

when $w = 0.03$, $L = 2.272$, and $\delta = 0.5$, ARL_1 values were 20.05 and 3.43 (in Tables 9.1 and 9.5)

for $\rho = 0.05$ and $\rho = 0.95$, respectively. Thus, increases in the correlation structure between the process variable and the auxiliary variable leads to an increase in the chart's ability to detect a shift.

- The chart is ARL unbiased. That is, the ARL_1 values never exceed the corresponding ARL_0 for any choice of δ examined.
- As δ increases, the ARL_1 and $SDRL_1$ values approach 1 and 0, respectively, especially for large values of ρ ; that is, the charts detect large shifts promptly.

Comparisons

We provide detailed comparisons of the proposed AHWMA chart with some existing control charts: the classical HWMA chart by Abbas (2018), the classical EWMA chart by Roberts (1959), the classical CUSUM chart by Roberts (1959), the auxiliary-based EWMA chart (i.e., M_XEWMA) by Abbas et al. (2014b), and the auxiliary-based CUSUM chart (i.e., $A_{ux}CUSUM_2$ by Ridwan A. Sanusi (2017), in terms of their ARL values. The auxiliary-based EWMA and CUSUM charts provide efficient applications of the classical EWMA and CUSUM charts, respectively, in those situations where the process variable is observed along with an variable; they are also based on a regression estimator. For comparison with the M_XEWMA and $A_{ux}CUSUM_2$ charts, we considered three different values of ρ : namely, $\rho \in \{0.05, 0.5, 0.95\}$. In all cases, the charts' parameters were set to values that fix ARL_0 at 500. We provide the charts' ARL results that optimized δ at $w \in \{0.05, 0.1, 0.2\}$.

The results of the comparisons are provided in Table 9.6. As shown on the table, the AHWMA chart outperformed the classical CUSUM, EWMA and HWMA charts in detecting shifts in the mean, especially when $\rho > 0.05$. For fixed values of w and ρ , the AHWMA chart was more efficient than the $A_{ux}CUSUM_2$ chart, especially for small-to-moderate values of δ (i.e., $\delta < 2$). For fixed values of w and ρ , the AHWMA chart was less efficient than the M_XEWMA chart in detecting moderate-to-large shifts (i.e., $\delta > 0.5$) in

Table 9.6: ARL comparisons of the charts.

		Classical chart			A _{ux} CUSUM ₂				M _X EWMA				AHWMA			
					ρ				ρ				ρ			
w	δ	CUSUM	EWMA	HWMA	0.05	0.5	0.75	0.95	0.05	0.5	0.75	0.95	0.05	0.5	0.75	0.95
0.03	0	497.95	500.3	501.2	498.9	497.43	501.98	501.22	502.7	500.82	503.52	498.71	502.98	502.09	501.65	502.44
	0.050	399.33	389.49	365.09	400.21	374.31	319.5	149.88	388.63	362.25	304.3	131.56	363.45	333.66	271.72	115.73
	0.075	320.31	304.9	273.63	322.98	287.25	222.65	88.09	303.73	271.43	205.42	70.58	273.65	240.95	178.17	65.13
	0.100	254.31	233.62	205.84	252.59	219.7	161.3	60.28	236.25	201.31	143.61	44.75	206.62	174.83	125.55	41.61
	0.125	201.09	181.26	159.07	200.39	170.07	121.51	45.42	182.06	151.38	103.07	30.82	159.27	132.93	91.66	28.84
	0.150	161.81	143.7	125.81	162.17	135.08	95	36.36	144.63	117.7	77.65	22.71	126.43	104.06	70.54	21.43
	0.175	133.45	116.37	102.08	133.18	110.83	77.63	30.2	114.6	93.03	60.51	17.48	102.59	83.71	55.94	16.54
	0.200	111.78	94.55	84.62	112.05	92.38	65.02	25.94	94.25	75.25	49.1	14.07	84.61	69.32	45.3	13.35
	0.250	83.19	66.23	61.21	83.33	69.13	48.66	20.16	66.71	52.69	33.91	9.66	60.59	49.08	31.95	9.32
	0.500	34.63	21.26	20.01	34.65	29.26	21.5	9.59	21.42	16.69	10.62	3.12	20.05	15.75	10.24	3.43
	0.750	21.67	10.75	10.29	21.62	18.45	13.76	6.37	10.71	8.44	5.4	1.75	10.31	8.23	5.5	1.99
	1.000	15.75	6.63	6.61	15.74	13.5	10.16	4.82	6.63	5.23	3.4	1.26	6.57	5.35	3.69	1.33
	1.500	10.27	3.44	3.72	10.24	8.85	6.75	3.3	3.44	2.78	1.88	1.01	3.74	3.1	2.16	1.01
	2.000	7.64	2.25	2.55	7.64	6.62	5.1	2.67	2.23	1.85	1.33	1	2.55	2.1	1.45	1
0.05	0	499.43	497.84	497.55	498.96	497.14	500.21	497.54	499.34	499.04	496.68	499.39	498.7	498.36	500.95	499.63
	0.050	425.38	412.14	380.67	420.29	404.36	355.01	175.78	410.96	388.75	334.74	154.71	382.76	354.43	297.52	133.96
	0.075	355.89	336.92	297.09	355.33	326.01	260.3	97.17	335.44	302.76	236.23	83.31	295.14	263.31	201.34	77.2
	0.100	291.22	266.23	230.76	289.35	253.87	187.5	61.51	267.56	231.69	168.05	51.09	228.92	197.66	144.11	50.65
	0.125	234.83	213.89	180.59	235.22	198.79	139.53	43.36	210.55	177.37	121.77	34.75	180.75	151.91	108.24	35.68
	0.150	189.47	168.75	146.07	190.51	157.49	106.35	33.11	168.53	137.9	91.73	25.44	144.81	120.7	84	26.85
	0.175	154.72	136.38	119.93	154.38	126.25	83.45	26.58	134.81	108.29	70.33	19.44	119.59	98.45	67.61	20.62
	0.200	128.86	110.55	100.46	127.41	103.56	67.15	22.14	110.86	88.48	56.2	15.57	100.41	81.65	54.88	16.58
	0.250	91.21	77.73	73.05	90.94	72.8	47.31	16.57	77.52	60.93	38.44	10.65	72.93	59.23	39.23	11.5
	0.500	31.14	23.73	24.86	31.28	25.52	17.83	7.39	23.73	18.45	11.72	3.4	24.89	19.73	12.66	4.01
	0.750	18.06	11.91	12.81	17.9	15.03	10.9	4.83	11.83	9.31	5.96	1.87	12.74	10.12	6.68	2.33
	1.000	12.63	7.28	8.06	12.59	10.66	7.87	3.65	7.27	5.75	3.73	1.32	8.01	6.44	4.35	1.52
	1.500	7.94	3.77	4.42	7.93	6.78	5.12	2.52	3.76	3	2.03	1.02	4.42	3.61	2.52	1.02
	2.000	5.85	2.42	2.98	5.82	5.03	3.86	2.01	2.43	1.98	1.4	1	2.98	2.43	1.65	1
0.1	0	501.58	501.67	500.39	501.92	502.21	501.54	498.63	500.99	497.33	499.62	502.92	502.95	501.47	499.57	501.08
	0.050	471.36	437.93	397.64	469.78	461.18	437.92	298.48	439.33	420.52	377.94	199.61	396.35	373.92	315.26	148.56
	0.075	438.67	380.9	318.89	438.02	422.52	374.37	190.06	377.32	350.23	287.74	110.71	317.18	285.65	222	86.73
	0.100	397.64	318.94	249.81	398.67	372.84	311.88	119.6	318.25	281.56	213.89	66.65	250.84	216.45	160.49	56.68
	0.125	354.81	262.15	200.21	356.1	320.97	253.42	78.24	260.83	227.52	159.78	43.91	198.63	168.43	120.11	40.43
	0.150	312.61	215.56	161.03	315.18	277.04	204.92	53.32	216.54	181.54	121.34	30.94	162.37	135.03	93.52	30.69
	0.175	273.9	178.62	133.14	273.86	233.99	164.5	37.95	177.41	145.27	94.14	23.26	132.35	110.08	75.41	23.92
	0.200	237.27	146.13	111.78	235.96	200.4	132.38	28.07	147.45	118.36	73.99	18.16	111.43	91.18	61.84	19.31
	0.250	179.15	102.97	81.64	178.76	143	88.48	17.08	103.48	80.95	48.77	12.15	81.49	65.89	44.38	13.41
	0.500	48.81	28.7	28.51	48.63	35.23	19.31	5.01	28.64	21.95	13.41	3.75	28.54	22.63	14.71	4.5
	0.750	19.62	13.6	14.87	19.66	14.55	8.64	2.95	13.59	10.52	6.59	2.03	14.88	11.79	7.64	2.59
	1.000	10.93	8.21	9.34	10.93	8.43	5.43	2.17	8.25	6.42	4.14	1.4	9.33	7.43	4.91	1.69
	1.500	5.5	4.17	4.96	5.47	4.48	3.15	1.46	4.15	3.32	2.22	1.02	4.96	4.04	2.8	1.04
	2.000	3.69	2.65	3.32	3.67	3.09	2.3	1.04	2.65	2.15	1.5	1	3.31	2.72	1.86	1
0.25	0	503.87	501.19	498.41	499.91	498.65	496.72	499.94	499.65	500.9	497.9	497.83	501.41	504.88	502.97	501.37
	0.050	473.65	464.89	440.22	476.76	467.82	443.37	316.9	469.38	454.7	429.18	284.8	438.68	425.24	385.37	209.06
	0.075	447.21	433.74	382.85	440.66	430.85	390.87	211.47	433.41	413.71	368.59	179.31	384.78	354.88	296.36	120.3
	0.100	411.5	390.42	324.23	412.17	387.72	334.12	138.29	387.97	361.12	298.88	114.37	326.26	290.14	223.57	74.83
	0.125	374.92	344.94	271.63	372.42	343.61	277.01	91.75	342.65	310.56	241.79	75.25	272.98	234.95	171.01	50.99
	0.150	335.01	300.26	224.11	333.73	297.96	228.79	62.32	301.25	265.16	194.2	51.62	226.05	190.51	131.78	36.25
	0.175	296.57	263.35	187.63	295.29	258.58	184.72	44.02	260.14	224.89	156.58	37.1	185.95	153.84	103.9	27.41
	0.200	260.28	225.42	157.86	261.6	221.65	152.31	31.91	226.33	190.14	126.5	27.39	156.47	127.81	83.18	21.51
	0.250	199.46	168.99	113.14	200.01	162.65	101.91	19.06	169.1	136.54	83.92	16.88	112.27	89.86	55.98	14.43
	0.500	57.14	47.4	33.79	56.76	40.9	21.68	4.89	47.02	34.6	18.99	4.24	33.81	25.77	15.96	4.45
	0.750	22.16	19.34	16.25	21.95	15.9	8.94	2.81	19.19	14.13	8.1	2.22	16.13	12.52	7.79	2.46
	1.000	11.57	10.45	9.71	11.56	8.69	5.36	2.04	10.39	7.77	4.73	1.49	9.67	7.59	4.86	1.62
	1.500	5.42	4.77	4.93	5.42	4.34	3	1.3	4.75	3.72	2.42	1.04	4.93	3.95	2.67	1.04
	2.000	3.55	2.93	3.18	3.54	2.93	2.16	1.02	2.94	2.34	1.6	1	3.19	2.61	1.76	1
	C	13.15029	2.483	2.272	13.15029	13.15029	13.15029	13.15029	2.483	2.483	2.483	2.483	2.272	2.272	2.272	2.272

the process mean. However, the chart shows greater efficiency than the M_XEWMA chart in detecting small shifts (i.e, $\delta \leq 0.5$) in the mean.

9.4 Robustness to non-normality of the chart

The AHWMA chart described in Section 9.2 relies on the assumption that the process variable and the auxiliary variable are bivariate normally distributed. In practice, this assumption does not always hold. Non-normality is not a major concern with a large sample size because the central limit theorem warrants that the sample mean will be approximately normally distributed for any continuous variables (Stoumbos and Sullivan, 2002). When $n = 1$, however, it is important to check the sensitivities of control charts to departures from normality (Testik et al., 2003). We refer readers to Stoumbos and Reynolds (2000); Maravelakis et al. (2005); Human et al. (2011), and Borrór et al. (2018) for detailed studies on the robustness of the EWMA control chart to non-normality.

Here, we investigate the robustness of the AHWMA chart to non-normality. As mentioned by Human et al. (2011) “a control chart is robust if its in-control run-length distribution remains stable (unchanged or nearly unchanged) when the underlying distributional assumption(s) (e.g. normality) are violated”. Following previous investigators (Stoumbos and Reynolds, 2000; Maravelakis et al., 2005; Human et al., 2011; Borrór et al., 2018; Aslam et al., 2019), we considered a heavy-tailed bivariate distribution - the bivariate Student’s t -distribution, and a skewed distribution - the bivariate gamma distribution. We denote the bivariate t -distribution with v degrees of freedom by $t_2(v)$. The probability density function of a bivariate t -distribution is given by

$$f(\mathbf{x}) = \frac{[\Gamma(v+2)/2]}{\Gamma(v/2)v\pi|\Sigma|^{1/2}} \left[1 + \frac{1}{v}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu}) \right]^{-(v+2)/2} \quad (9.9)$$

where $\mathbf{x} \in R^2$, $\boldsymbol{\mu} = [\mu_1, \mu_2]^T$ is the 2×1 vector of location parameters, Σ is a 2×2 positive-definite (or covariance) matrix, v is the number of the degrees of freedom, and $\Gamma(n) = (n-1)!$ for $n = 1, 2, \dots$. The mean vector and covariance matrix are given as $\boldsymbol{\mu}$ (if $v > 1$, else, undefined), and $\frac{v}{(v-2)}\Sigma$ (when $v > 2$, else, undefined), respectively.

We denote the bivariate gamma distribution with shape parameter, $\alpha \mathbf{1}_p$, and scale parameter, $\beta \mathbf{1}_p$, by $G_2(\alpha \mathbf{1}_p, \beta \mathbf{1}_p, \Sigma)$, where $\mathbf{1}_p$ is a column vector of ones of size $p = 2$. The probability density function of the bivariate gamma distribution is given as:

$$f(x) = \frac{|\Sigma|^{-\alpha}}{\beta^{2\alpha} \Gamma_2(\alpha)} |\mathbf{x}|^{\alpha-3/2} \exp \left(tr \left(\frac{1}{\beta} \Sigma^{-1} \mathbf{x} \right) \right) \quad (9.10)$$

where $\alpha > 0$ is the scale parameter, $\beta > 0$ is the shape parameter, and Γ_2 is the bivariate gamma function given as $\Gamma_2(\alpha) = \pi^{1/2} \Gamma(\alpha) \Gamma(\alpha - 1/2)$. See Ronning (1977) and Gupta, Arjun K and Nagar (2018) for detailed information on the bivariate gamma distribution and its properties.

We studied the chart's robustness under a large range of degrees of freedom (v) for the bivariate t -distribution; namely, $v \in \{4, 6, 8, 10, 15, 20, 30, 40, 50, 100, 1000\}$. For the bivariate gamma distribution, without loss of generality, we considered scale parameter, $\beta = 1$, and a range of values of the shape parameter, i.e., $\alpha \in \{1, 2, 3, 4, 5, 10, 50, 100, 1000\}$. Hence, we denote the bivariate gamma distribution as $G_2(\alpha)$, for short. The ARL_0 values of the chart for $\rho \in \{0.25, 0.5, 0.95\}$, and $w \in \{0.03, 0.05, 0.25, 0.75\}$ for the bivariate t and bivariate gamma distributions are given in Tables 9.7 and 9.8, respectively.

Table 9.7: ARL_0 with bivariate t -distribution

w	ρ_{ZY}	AHWMA	$t_2(4)$	$t_2(8)$	$t_2(10)$	$t_2(15)$	$t_2(30)$	$t_2(50)$	$t_2(100)$	$t_2(1000)$
0.03	0.25	502.5	268.66	346.71	372.71	413.4	454.92	472.15	485.14	499.74
	0.5	502.09	268.16	346.42	372.37	413.24	454.77	472.99	485.93	499.36
	0.95	502.44	268.4	346.48	372.04	413.12	454.27	472.85	485.4	499.98
0.05	0.25	498.92	223.83	306.24	332.44	375.3	433.23	454.27	476.72	494.77
	0.5	498.36	223.1	305.14	332.2	375.7	432.25	455.71	477.37	494.7
	0.95	498.97	223.74	305.58	332.1	375.69	432.53	455.09	477.4	494.65
0.25	0.25	504.01	103.84	163.35	188.54	237.89	328.6	381.05	433.02	489.41
	0.5	504.88	103.87	163.21	188.52	237.59	328.94	381.32	433.3	489.25
	0.95	504.97	104.53	163.6	188.42	237.36	328.48	381.82	433.04	489.66
0.75	0.25	497.98	84.55	137.14	163.15	213.72	307.88	366.82	429.93	485.26
	0.5	496.55	84.37	137.01	163.27	213.11	307.99	366.91	429.73	485.26
	0.95	497.6	84.29	137.96	163	213.55	307.67	366.43	429.51	485.68

The ARL_0 results in Tables 9.7 and 9.8 are summarized below:

Table 9.8: ARL₀ with bivariate gamma distribution

w	ρ_{ZY}	AHWMA	$G(1)$	$G(2)$	$G(3)$	$G(4)$	$G(5)$	$G(50)$	$G(100)$	$G(1000)$
0.03	0.25	502.5	318.96	375.58	403.3	423	429.81	490.83	499.57	501.11
	0.5	502.09	276.22	331.92	366.38	388.89	402.29	486.64	495.02	499.28
	0.95	502.44	203.56	260.01	301	326.33	345.62	474.14	486.89	499.52
0.05	0.25	498.92	260.97	316.67	348.37	377.24	391.84	484.87	496.06	498.53
	0.5	498.36	228.92	281.5	321.79	345.27	366.05	482.82	482.17	496.23
	0.95	498.97	169.00	222.22	260.75	285.58	308.43	457.36	480.23	495.58
0.25	0.25	504.01	91.82	120.34	139.28	159.94	174.75	394.77	444.34	492.63
	0.5	504.88	87.9	117.06	139.89	159.19	175.89	398.25	438.8	492.79
	0.95	504.97	65.44	95.52	117.63	139.88	160.81	386.1	436.85	485.02
0.75	0.25	497.98	61.61	82.27	98.38	115.33	127.7	352.01	410.57	491.43
	0.5	496.55	59.7	82.53	96.01	114.57	125.31	351.67	409.13	484.35
	0.95	497.6	46.52	71.75	92.58	113.61	122.13	344.77	410.59	480.14

- For a fixed value of w , the ARL₀ values of the bivariate t -distributions are the same for all the correlation values (i.e., $\rho_{ZY} = 0.25, 0.5$, or 0.95) examined. This result is due to the symmetric nature of the t -distribution.
- However, for the bivariate gamma distributions, for a fixed value of w , the ARL₀ differ across all the values of ρ_{ZY} examined. Here, the chart appears to be more robust to non-normality only for smaller values of ρ_{ZY} .
- For both non-normal distributions, as expected, the ARL₀ value increases, and tends to converge to the required nominal ARL₀ of the AHWMA chart, for large degrees of freedom (v) or larger values of the shape parameter (i.e., $\alpha \geq 50$), especially when $w = 0.3$ or 0.05 is used.
- Importantly, the chart's ARL₀ value is more robust to non-normality only when a small value of w (i.e., $w = 0.03$ or 0.05) is used. This implies that small values of w (i.e., $w = 0.03$ and 0.05) are fairly useful when the underlying distribution is not normal.

Table 9.9 displays the ARL₁ values for the AHWMA chart under bivariate normal, t and gamma distributions for various values of δ when $w = 0.03$, or 0.75 , and $\rho_{ZY} = 0.25$. The results in Table 9.9

indicate that the chart's ARL_1 values tend to approach values obtained for bivariate normal data when a smaller value of w (i.e., $w = 0.03$) is used. For example, when $w = 0.03$, $v = 50$, $\beta = 50$, and $\delta = 0.5$, the ARL_1 for the AHWMA were 19.02 (normal distribution), 19.65 (t -distribution), and 19.18 (gamma distribution). The percentage deviation of the ARL_1 values obtained under the t or gamma distributions from ARL_1 values obtained under normal distribution are 3.13% and 0.84%, respectively. On the other hand, when w is large (i.e., $w = 0.75$), and other parameters are unchanged (i.e., $v = 50$, $\beta = 50$, and $\delta = 0.5$), then the ARL_1 for the AHWMA under normal, t and gamma distributions were 125.44, 118.63, and 68.10, respectively; the percentage deviation of these ARL_1 values from obtained under the normal distribution were -5.43% and -45.71% for the t and gamma distributions, respectively.

9.5 Step by step algorithm for constructing the AHWMA chart when parameters are unknown

The AHWMA chart in Section 9.2 was formulated assuming parameters associated with the process variable and auxiliary variable are all known; however, these parameters are generally unknown in practice and need to be estimated. In this case, the regression model in Equation (9.3) would be based on estimated parameters, and is given as:

$$\hat{R}_i = \bar{z}_i + \hat{b}(\hat{\mu}_Y - \bar{y}_i) \quad (9.11)$$

where \hat{b} is the estimated slope of the regression line; given as the estimated change in the process variable Z due to a unit change in the auxiliary variable Y (Cochran, 1977), and $\hat{\mu}_Y$ ($\hat{\mu}_Y = \frac{1}{m} \sum_{i=1}^m \bar{Y}_i$) is the unbiased estimate of the mean of the auxiliary variable (i.e., μ_Y). The estimated mean and variance of \hat{R} are given as $\bar{\hat{R}} = \hat{\mu}_Z$, and $S_{\hat{R}}^2 = \frac{\hat{\sigma}_Z^2}{n}(1 - r^2)$, where r is the estimated value of the correlation size between the variables, $\hat{\mu}_Z$ and $\hat{\sigma}_Z^2$ are the unbiased estimates of μ_Z and σ_Z^2 , respectively. The $\hat{\mu}_Z$ and

Table 9.9: ARL₁ with bivariate t and gamma Distributions

w	ρ		δ					
			0.1	0.2	0.25	0.5	1	2
0.03	0.25	AHWMA	199.08	81.18	58.11	19.02	3.57	2.45
		$t_2(4)$	208.67	121.2	92.89	33.63	10.79	3.92
		$t_2(8)$	207.97	95.66	71.22	24.01	7.83	2.97
		$t_2(10)$	206.5	92.86	68.27	22.93	7.47	2.83
		$t_2(15)$	204.56	88.88	63.92	21.42	7.05	2.71
		$t_2(30)$	202.67	84.32	61.18	20.23	6.64	2.57
		$t_2(50)$	201.41	83.31	59.57	19.65	6.54	2.53
		$t_2(100)$	201.33	82.01	59.37	19.39	6.44	2.49
		$t_2(1000)$	201	81.45	58.84	18.91	6.32	2.45
		$G(1, 1)$	145.21	71.65	54.11	19.54	7.07	2.98
		$G(2)$	154.77	73.77	54.64	19.53	6.79	2.84
		$G(3)$	162.67	74.64	55	19.39	6.63	2.77
		$G(4)$	163.66	75.48	56.17	19.33	6.62	2.72
		$G(5)$	169.73	75.68	56.08	19.23	6.62	2.7
		$G(50)$	186.67	80.17	57.93	19.18	6.39	2.55
		$G(100)$	187.92	80.29	57.8	19.2	6.36	2.49
		$G(1000)$	196.47	80.52	58.36	19.09	6.39	2.46
0.75	0.25	AHWMA	457.88	356.45	305	125.44	8.38	3.93
		$t_2(4)$	84.1	81.67	80.5	71.4	46.83	13.02
		$t_2(8)$	134.59	125.99	121.83	87.94	35.57	6.16
		$t_2(10)$	157.11	147.49	140.22	92.84	32.98	5.48
		$t_2(15)$	205.46	187.12	171.97	103.76	30.63	4.85
		$t_2(30)$	291.42	246.38	224.29	114.43	28.34	4.28
		$t_2(50)$	344.22	285.82	251.86	118.63	27.08	4.17
		$t_2(100)$	391.79	312.18	274.8	120.82	26.45	4.05
		$t_2(1000)$	448.32	354.74	299.77	124.34	25.83	3.89
		$G(1)$	47.75	38.57	34.32	21.5	9.77	3.68
		$G(2)$	63.26	49.67	44.14	26.01	11.04	3.68
		$G(3)$	75.12	57.82	52.45	29.26	11.86	3.7
		$G(4)$	87.47	66.01	57.66	32.31	12.48	3.7
		$G(5)$	95.62	72.83	63.7	34.49	13.01	3.66
		$G(50)$	250.86	177.39	150.4	68.1	18.63	3.77
		$G(100)$	303.73	215.39	179.48	79.03	20.29	3.8
		$G(1000)$	407.08	296.01	258.56	106.41	23.41	3.87

$\hat{\sigma}_Z^2$ are calculated from a specified set of sample values measured when the process was known to be in control, and are given as $\hat{\mu}_Z = \frac{1}{m} \sum_{i=1}^m \bar{Z}_i$, and $\hat{\sigma}_Z = \frac{s_p}{c_{4,m}}$, where $s_p = \left(\frac{\sum_{i=1}^m \sum_{j=1}^n (Z_{ij} - \bar{Z}_i)^2}{m(n-1)} \right)^{1/2}$,

and $c_{4,m} = \frac{2^{1/2}\Gamma\left(\frac{m(n-1)+1}{2}\right)}{(m(n-1))^{1/2}\left(\frac{m(n-1)}{2}\right)}$ is an un-biasing constant (Jones et al., 2001; Abbas, 2018).

Using Equation (9.11), the plotting statistic for the AHWMA control chart based on estimated parameters is given as:

$$\hat{T}_i = w\hat{R}_i + (1-w)\bar{\hat{R}}_{i-1} \quad (9.12)$$

The estimated mean and variance of the plotting statistic in Equation (9.12) are given as $\hat{\mu}_T = \hat{\mu}_Z$, and

$$\hat{\sigma}_{\hat{T}_i}^2 = \begin{cases} \frac{(1-r^2)}{n} w^2 \hat{\sigma}_Z^2 & \text{if } i = 1 \\ \frac{(1-r^2)}{n} \left(w^2 \hat{\sigma}_Z^2 + (1-w)^2 \frac{\hat{\sigma}_Z^2}{i-1} \right) & \text{if } i > 1 \end{cases} \quad (9.13)$$

The upper and lower control limits for the (plotting statistic given in Equation (9.12)) estimated time varying control chart are given as:

$$\hat{L}_i = \begin{cases} \hat{\mu}_Z - C' \hat{\sigma}_Z \sqrt{\frac{w^2}{n} (1-r^2)} & \text{if } i = 1 \\ \hat{\mu}_Z - C' \hat{\sigma}_Z \sqrt{\left(\frac{w^2}{n} + \frac{(1-w)^2}{n(i-1)} \right) (1-r^2)} & \text{if } i > 1 \end{cases} \quad (9.14)$$

$$\hat{U}_i = \begin{cases} \hat{\mu}_Z + C' \hat{\sigma}_Z \sqrt{\frac{w^2}{n} (1-r^2)} & \text{if } i = 1 \\ \hat{\mu}_Z + C' \hat{\sigma}_Z \sqrt{\left(\frac{w^2}{n} + \frac{(1-w)^2}{n(i-1)} \right) (1-r^2)} & \text{if } i > 1 \end{cases} \quad (9.15)$$

where C' determines the width of the estimated control limits. Also, the estimated centre line (CL) of the AHWMA chart is given by:

$$\hat{CL} = \hat{\mu}_Z \quad (9.16)$$

When the chart is based on estimated parameters, implementation occurs in two phases. In phase I

(retrospective phase), a historical reference sample is studied to establish the in-control state and to evaluate the stability of the process (Jensen et al., 2006; Abbasi and Adegoke, 2018). Once the in-control reference sample is characterised, the process parameters are estimated from phase I, and control chart limits are obtained for use in phase II. The phase II aspect initiates ongoing regular monitoring of the process. If successive observed values obtained at the beginning of Phase II fall within the in-control limits calculated from Phase I, the process is considered to be in control. In contrast, any observed values during Phase II which fall outside the control limits indicate that the process may be out of control, and remedial responses are then required (Montgomery, 2009; Adegoke et al., 2018b). A shift in a process parameter needs to be detected quickly so that corrective actions can be taken at an early stage.

We give below a step-by-step algorithm to implement chart in phase I and phase II (Abbas, 2018; Mahmouda and Maravelakisb, 2010).

- Phase I

1. Simulate m bivariate samples (Z, Y) each of size n from the in-control historical process.
2. Calculate the sample means, (\bar{Z}_i, \bar{Y}_i) , and the sample variances (S_{Zi}^2, S_{Yi}^2) , where $S_{Zi}^2 = \frac{\sum_{j=1}^n (Z_{ij} - \bar{Z}_i)^2}{(n-1)}$, and $S_{Yi}^2 = \frac{\sum_{j=1}^n (Y_{ij} - \bar{Y}_i)^2}{(n-1)}$, for each sample $i = 1, 2, \dots, m$.
3. Repeat steps 1 and 2, many times, and compute estimates for the means, $(\hat{\mu}_Z, \hat{\mu}_Y)$ and variances $(\hat{\sigma}_Z^2, \hat{\sigma}_Y^2)$. These are used to set the control limits in phase II.

- Phase II

4. At each time i , simulate a bivariate samples of size n from the process.
5. Compute the estimated regression estimator in Equation (9.11), and use this to compute the chart's plotting statistic, \hat{T}_i , in Equation (9.12).
6. Use the estimated parameters from phase I (from step 3 above), and construct the estimated

control limits given in Equation (9.14) through Equation (9.15). Compared \hat{T}_i , against these control limits.

7. If \hat{T}_i falls within the control limits, the process is declared to be in control. Alternatively, if \hat{T}_i falls outside the control limits, the process is declared to have shifted to an out-of-control state.

9.6 Industrial application

In this section, we provide an illustrative example to show an application of the AHWMA chart, using a dataset from a similar study by Abbas et al. (2014b). The data were obtained by simulating $m = 20$ samples each of size $n = 1$ from $(Z, Y) \sim N_2(\mu_Z + \delta\sigma_Z, \mu_Y, \sigma_Z^2, \sigma_Y^2, \rho)$. The values of the parameters used for the simulation were: $\mu_Z = 0$, $\mu_Y = 0$, $\sigma_Z^2 = 1$, $\sigma_Y^2 = 1$, $\rho = 0.5$, and $\delta = 0.5$, where δ is the size of the shift applied to the in-control mean, μ_Z , of the process variable of interest, and ρ is the size of the correlation between the process variable and the auxiliary variable. We examined the ability of the AHWMA chart to detect a shift in the process variable and compared this to the M_X EWMA chart, as well as the classical EWMA and HWMA charts. In all cases, the chart parameters: w and C , were chosen to fix ARL_0 to 500. The parameters for the classical EWMA and M_X EWMA were $w = 0.03$ and $C = 2.483$ (see Table 9.6); for the classical HWMA and AHWMA charts, we used $w = 0.03$ and $C = 2.272$ (see Table 9.6). We give the calculations for the AHWMA chart in Table 9.10, and the results for all the control charts are shown graphically in Figure 9.1.

The AHWMA chart detected the shift in the process mean faster than any of the other methods. In particular, it detected the shift after the 14th sample, whereas the M_X EWMA chart detected the shift after the 15th sample, and the classical EWMA, and HWMA charts both detected the shift only after the 18th sample.

Table 9.10: Calculation of the AHWMA chart statistic and its limits

i	Z_i	Y_i	R_i	T_i	LCL_i	UCL_i
1	0.39	-0.865	0.8225	0.0247	-0.059	0.059
2	-0.242	-1.686	0.601	0.8159	-1.9095	1.9095
3	-0.919	-1.046	-0.396	0.6785	-1.3509	1.3509
4	-1.22	-1.366	-0.537	0.3161	-1.1035	1.1035
5	2.01	0.574	1.723	0.1706	-0.9561	0.9561
6	1.395	1.61	0.59	0.4471	-0.8556	0.8556
7	1.66	1.542	0.889	0.4799	-0.7814	0.7814
8	-0.514	0.816	-0.922	0.484	-0.7238	0.7238
9	-0.213	-0.907	0.2405	0.3431	-0.6774	0.6774
10	-0.588	-1.923	0.3735	0.3357	-0.6389	0.6389
11	0.074	0.132	0.008	0.3285	-0.6064	0.6064
12	1.673	1.64	0.853	0.3247	-0.5785	0.5785
13	1.765	0.575	1.4775	0.3875	-0.5541	0.5541
14	0.061	-0.008	0.065	0.429	-0.5326	0.5326
15	1.537	-1.084	2.079	0.4634	-0.5135	0.5135
16	-0.519	-0.52	-0.259	0.5010*	-0.4963	0.4963
17	1.198	-0.246	1.321	0.5009*	-0.4808	0.4808
18	1.853	0.028	1.839	0.5646*	-0.4666	0.4666
19	0.733	1.715	-0.1245	0.5765*	-0.4537	0.4537
20	0.108	-0.6	0.408	0.5556*	-0.4418	0.4418

* Out-of-control signal

9.7 Conclusion and Discussion

We propose here a new efficient control-chart method, for monitoring small shifts in the process mean where the process variable of interest is correlated with and observed alongside an auxiliary variable. Based on the homogeneously weighted moving average, the proposed chart uses both the process and auxiliary variable to form a regression estimator that yields an efficient and unbiased estimate of the mean of the process variable. We provided the design structure of the chart and examined its performance in terms of its run length properties. Our simulation results showed that the chart detects a shift in the process mean more rapidly than other methods. Also, the ARL comparisons showed that the chart we propose here is generally more efficient than existing control charts used for the same purpose, especially when interest lies in detecting a small shift in the process mean. We provided a detailed study of the chart's robustness

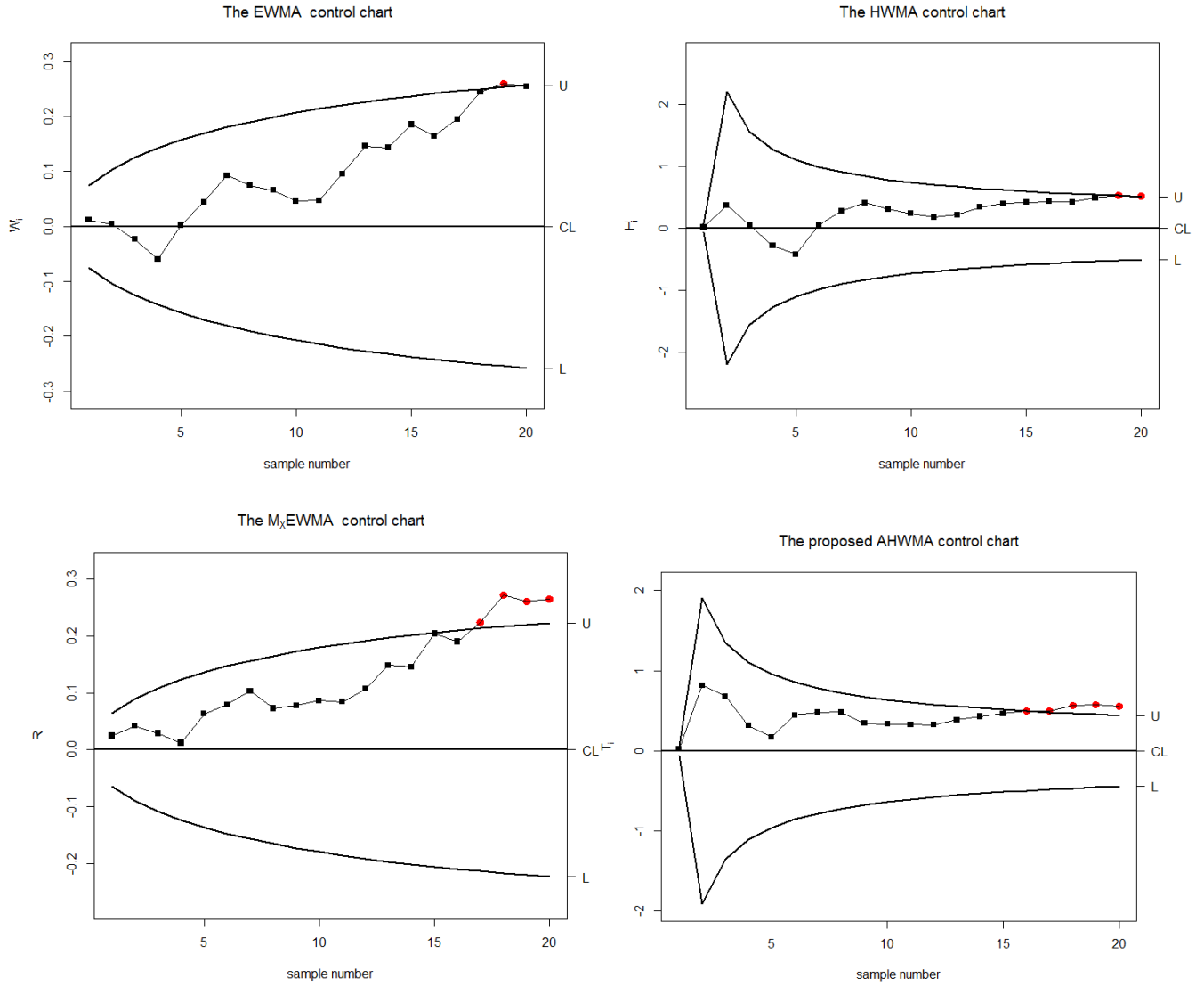


Figure 9.1: Application of the EWMA, HWMA, M_X EWMA, and AHWMA charts.

to non-normality and showed how it can be designed to reduce sensitive to non-normality. The chart's ARL values showed that the chart is more robust to non-normality when a smaller value of w is used. In particular, when a small value is chosen for the chart's smoothing parameter (for example $w \leq 0.05$), the proposed chart can be designed to have an in-control ARL that is reasonably close to the ARL for the chart under a normally distributed process. We gave some recommendations on the application of the chart when the process parameters are unknown, and provided a step-by-step algorithm to construct the chart for phase I and phase II of SPC. Also, we applied the chart to a simulated dataset and showed

that it detected a small shift in the process mean faster than other examined charts including EWMA, HWMA, M_X EWMA, and A_{ux} CUSUM₂ methods. We consider that the effect of estimating parameters during phase I of the process on subsequent performance of the AHWMA chart warrants further study.

Chapter 10

Summary and Recommendations for Future Research

10.1 Summary

This doctoral thesis develops and improves control chart methods for applications in statistical, ecological, and industrial fields by proposing new methods and also by improving currently available methods for efficient detection of (unusual) signals in monitored systems.

The study presented in chapter 2 demonstrates that, if the mean vector and covariance matrix parameters of the multivariate cumulative sum (MCUSUM) and the multivariate CUSUM #1 (MCI) control charts for individual-observation sample size (i.e., $n = 1$) are unknown, the in-control performance of the charts are negatively affected by estimating the process parameters during Phase I, unless a large number of Phase I observations ($m > 300$) are available. Control charts obtained using a shrinkage estimate of the covariance matrix along with an empirical estimate of the mean vector were found to be superior (in terms of run-length properties) to those obtained using other estimation methods. The study also

provided the values of the corrected control limits that will give a desired in-control ARL performance when parameters are estimated, along with a least-squares model that can be used to estimate these corrected limits for new situations. The relative performance of the control charting schemes, including the superiority of the shrinkage estimation method, was also readily illustrated in a particular example - the analysis of a multivariate Bimetal dataset.

The study in chapter 3 showed that the in-control performance of the multivariate exponentially weighted moving average (MEWMA) control chart is also strongly affected by the estimation procedure used, unless the Phase I sample is large. The findings from this study mainly showed that MEWMA control charts which used the shrinkage estimate of the covariance matrix consistently performed better than other methods, especially when only a small number of Phase I samples (i.e., $m \leq 300$) were available. On average, the shrinkage estimate had longer ARL_0 for in-control processes and shorter ARL for out-of-control processes. The study demonstrated superiority of the shrinkage estimation method with two real datasets. In both cases, the shrinkage-based chart detected true shifts earlier than the alternative methods. The shrinkage method also has the advantage of being applicable for high-dimensional data, in contrast with alternative methods, which cannot be used when the number of variables (p) exceeds the number of Phase I samples m . We recommend the use of the shrinkage estimate of the covariance matrix in situations where the parameter of the covariance matrix used in the development of the MEWMA, MCUSUM or MCI charts are unknown.

The study in chapter 4 showed that, for all the multivariate coefficient of variation (CV) charts examined in the chapter, the probability to signal (PTS) increases with an increase in sample size and the amount of shift in the multivariate CV. The findings from the study showed that the multivariate CV chart based on quadratic form (i.e., the MCV_{AZ} chart) exhibited the worst performance under both diffuse symmetric and localized CV disturbances. The multivariate CV chart based on a generalized Mahalanobis distance (i.e., the MCV_{VN}) chart outperformed other competing charts in the presence of

diffuse symmetric CV disturbances. The multivariate CV chart based on the geometric average of the eigenvalues of the covariance structure of the process (i.e., the MCV_R chart) performed very close to the MCV_{VN} chart and significantly better than the MCV_{AZ} chart. The MCV_R chart appeared to be the best choice in the presence of localized CV disturbances, as compared to either the MCV_{VN} or the MCV_{AZ} charts. The superiority of the MCV_R chart was also demonstrated with a real-life example using inner diameter and length of carbon tubes. We recommend the use of the MCV_R chart in Phase I of SPC, due to the fact that this chart performed the best for the detection of localized CV disturbances and its performance was very close to the MCV_{VN} chart in detecting diffuse symmetric disturbances.

Chapter 5 presented a new multivariate chart, namely, the multivariate homogeneously weighted moving average (MHWMA) control chart, for efficient monitoring of small shifts in the process mean vector. The performance of the chart was evaluated and compared with multivariate χ^2 , MEWMA, MCI and MCUSUM charts, across a variety of different possible charting parameters. The findings revealed that the proposed MHWMA chart is superior to the multivariate χ^2 , MEWMA, MCI and MCUSUM charts, particularly for the detection of small shifts in the process mean vector. Thus, we recommend the use of the MHWMA chart when interest lies in monitoring small shifts in the process mean vector.

Chapter 6 presented a new multivariate a distribution-free control chart tool for ecological monitoring. The proposed charting method does not require prior knowledge of the ecosystem's behavior before the monitoring begins, and was based on a change-point method where the currently observed ecological community was evaluated relative to previous measures of the community at a given site. The method was designed to take either the original multivariate dataset in Euclidean space, or the dissimilarities between every possible pair of samples using a dissimilarity measure of interest, as input. A permutation procedure was employed to obtain the control-chart limits for a suitable distance-based model through time. The findings showed that the method performed better than a previously described technique that uses distance to centroid of the multivariate ecological dataset, especially when there are association

among species variables in the ecological dataset. Examples were provided to show the application of the new proposed chart in detecting real shifts in ecological community structure.

Chapter 7 described the methods to enhance classical EWMA control charts for scenarios where the process variable is observed and is negatively correlated with an auxiliary variable under different sampling schemes: simple random sampling (SRS), ranked set sampling (RSS), and median ranked set sampling (MRSS). The findings from the study showed that the proposed charts are more efficient than the classical EWMA chart in detecting shifts in the mean of a process where the process variable is negatively correlated with the auxiliary variable, for some range of correlations (ρ_{XY}). In particular, the efficiency of the proposed charts over the classical EWMA control chart was more pronounced when there was a large negative correlation (i.e., $\rho_{XY} \leq -0.5$) between the variables, and these charts were also very efficient in detecting small shifts. Also, the findings showed that the proposed charts based on the MRSS, followed by the RSS, were more efficient than the proposed charts based on SRS. The proposed charts based on the MRSS or RSS are more applicable to those scenarios where measuring the process variable is expensive, but ranking of units according to the relative size of the process variable is easy. The performance of the proposed control charts was also illustrated with examples.

Chapter 8 extended the findings in chapter 7 to a more general correlation structure (positive or negative) between the process variable and the auxiliary variable. The performance of the proposed S_X EWMA and T_X EWMA charts based on the RSS and MRSS, respectively, were compared with available results (in the literature) based on simple random sampling (i.e., M_X EWMA). The findings showed that the proposed charts were more efficient than the classical EWMA and the M_X EWMA charts in detecting a shift in the location parameter of the process variable, even when there was only a small to moderate correlation (i.e., $|\rho_{XY}| \leq -0.5$) between the process variable and the auxiliary variable. Also, the findings showed that the proposed T_X EWMA chart was more efficient than the proposed S_X EWMA chart, especially for small to moderate shifts in the process mean. Moreover, the proposed S_X EWMA

and T_XEWMA charts performed better than the proposed charts based on RSS or MRSS in chapter 7, respectively.

Chapter 9 proposed a more efficient control chart for monitoring small shift in the process mean under simple random sampling (SRS), under scenarios where the process variable is correlated and observed with an auxiliary variable. The ARL comparison of the chart with several existing charts showed that the chart is more efficient than its competitors, especially when interest lies in detecting a small shift in the process mean. We provided a detailed study of the chart's robustness to non-normal distributions and showed that the chart also can be designed to be insensitive to non-normality. The chart's ARL values showed that the chart is more robust to non-normality when a smaller value of w is used. In particular, in the case of non-normality, when a small value of the chart's smoothing parameter is used (for example $w \leq 0.05$), the proposed chart can be designed to have an in-control ARL that is reasonably close to the ARL for the chart under a normally distributed process. We gave some recommendations on the application of the chart when the process parameters are unknown, and provided step-by-step algorithms to construct the chart in Phase I and Phase II of SPC. Also, application of the chart to a dataset showed that the proposed chart detected a small shift in the process mean faster than the other charts it was compared to. When interest lies in detecting small shift in the mean of a process that is (positively or negatively) correlated with another variable, under SRS, we recommend the proposed chart in this chapter because it outperformed the proposed charts in Chapters 7 and 8.

In general, the methods developed and examined in the thesis will aid practitioners in applying control charts for efficient monitoring.

10.2 Future Work

We consider that the following topics deserve further thought and directed research.

- The inertia property and the robustness to non-normality of the MHWMA chart in chapter 5 needs to be investigated. Also, the effect of parameter estimates on the Phase II performance of the MHWMA chart needs to be investigated.
- The proposed control charts in chapters 7 - 8, are only based on simple random sampling, ranked set sampling or median ranked set sampling. We recommend the proposed schemes be extended to other sampling schemes, such as double-stage sampling. Double-stage sampling involves applying the ranked set sampling and median ranked set sampling at different stages of monitoring.
- Extensions of the proposed chart in chapter 9 to ranked-based sampling schemes require further investigation. Also, the effect of estimating parameters on the AHWMA chart (in chapter 9) needs further study.

Bibliography

- Abbas, N. (2018). Homogeneously weighted moving average control chart with an application in substrate manufacturing process. *Computers and Industrial Engineering*, 120(May):460–470.
- Abbas, N., Raji, I. A., Riaz, M., and Al-Ghamdi, K. (2018). On Designing Mixed EWMA Dual-CUSUM Chart with Applications in Petro-Chemical Industry. *IEEE Access*, 6:78931–78946.
- Abbas, N., Riaz, M., and Does, R. J. (2013). Mixed exponentially weighted moving average-cumulative sum charts for process monitoring. *Quality and Reliability Engineering International*, 29(3):345–356.
- Abbas, N., Riaz, M., and Does, R. J. M. M. (2014a). An EWMA-Type Control Chart for Monitoring the Process Mean Using Auxiliary Information. *Communications in Statistics—Theory and Methods*, 43:3485–3498.
- Abbas, N., Riaz, M., and Does, R. J. M. M. (2014b). n EWMA-Type Control Chart for Monitoring the Process Mean Using Auxiliary Information An EWMA-Type Control Chart for Monitoring the Process Mean Using Auxiliary Information. *Communications in Statistics—Theory and Methods*, 43:3485–3498.
- Abbasi, S. A., Abbas, T., Riaz, M., and Gomaa, A. S. (2018). Bayesian Monitoring of Linear Profiles Using DEWMA Control Structures with Random X. *IEEE Access*, 6:78370–78385.
- Abbasi, S. A. and Adegoke, N. A. (2018). Multivariate coefficient of variation control charts in phase I of SPC. *International Journal of Advanced Manufacturing Technology*, 99(5-8):1903–1916.

- Abbasi, S. A., Miller, A., and Riaz, M. (2013). Nonparametric progressive mean control chart for monitoring process target. *Quality and Reliability Engineering International*, 29(7):1069–1080.
- Abbasi, S. A. and Riaz, M. (2013). On enhanced control charting for process monitoring. *International Journal of Physical Sciences Full Length Research Paper*, 8(17):759–775.
- Abbasi, S. A., Riaz, M., Miller, A., Ahmad, S., and Nazir, H. Z. (2015). EWMA Dispersion Control Charts for Normal and Non-normal Processes. *Quality and Reliability Engineering International*, 31(8):1691–1704.
- Abtew, M. A., Kropi, S., Hong, Y., and Pu, L. (2018). Implementation of Statistical Process Control (SPC) in the Sewing Section of Garment Industry for Quality Improvement. *Autex Research Journal*, 18(2):160–172.
- Abujiya, A. R. and Muttalak, H. A. (2004). Quality Control Chart for the Mean using Double Ranked Set Sampling. *Journal of Applied Statistics*, 31(10):1185–1201.
- Abujiya, Mu’azu Ramat and Farouk, Abbas Umar and Lee, Muhammad Hisyam and Mohamad, I. (2013). On the sensitivity of Poisson EWMA control chart. *International Journal of Humanities and Management Sciences*, 1:18—22.
- Adebola, FB and Adegoke, N. (2015). A Class of Regression Estimator with Cum-Dual Product Estimator as Intercept. *Global Journal of Science Frontier Research*, 15(3).
- Adebola, F.B. and Adegoke, N.A. and Sanusi, R. A. (2015). A Class of Regression Estimator with Cum-Dual Ratio Estimator as Intercept. *International Journal of Probability and Statistics*, 4(2):42—50.
- Adegoke, N. A., Abbasi, S. A., Dawod, A. B., and Pawley, M. D. (2018a). Enhancing the performance of the EWMA control chart for monitoring the process mean using auxiliary information. *Quality and Reliability Engineering International*, (October):1–14.

- Adegoke, N. A., Abbasi, S. A., Smith, A. N. H., Anderson, M. J., and Pawley, M. D. M. (2019). A Multivariate Homogeneously Weighted Moving Average Control Chart. *IEEE Access*, 7:9586–9597.
- Adegoke, N. A., Riaz, M., Sanusi, R. A., Smith, A. N., and Pawley, M. D. (2017). EWMA-type scheme for monitoring location parameter using auxiliary information. *Computers and Industrial Engineering*, 114:114–129.
- Adegoke, N. A., Smith, A. N., Anderson, M. J., Abbasi, S. A., and Pawley, M. D. (2018b). Shrinkage estimates of covariance matrices to improve the performance of multivariate cumulative sum control charts. *Computers and Industrial Engineering*, 117(February):207–216.
- Aerts, S., Haesbroeck, G., and Ruwet, C. (2015). Multivariate coefficients of variation: Comparison and influence functions. *Journal of Multivariate Analysis*, 142:183–198.
- Ahmad, S., Riaz, M., Abbasi, S. A., and Lin, Z. (2013). On monitoring process variability under double sampling scheme. *International Journal of Production Economics*, 142(2):388–400.
- Ajadi, J. O., Riaz, M., and Al-ghamdi, K. (2016). On increasing the sensitivity of mixed EWMA – CUSUM control charts for location parameter. *Journal of Applied Statistics*, 43(7):1262–1278.
- Al-Sabah, W. S. (2010). CUMULATIVE SUM STATISTICAL CONTROL CHARTS USING RANKED SET SAMPLING DATA. *Pak. J. Statist.*, 26(2):365–378.
- Alamand, Md Sarwar and Sinha, Arun Kumar and Ali, R. (2016). On Statistical Quality Control Techniques Based on Ranked Set Sampling. In *Recent Advances in Mathematics, Statistics and Computer Science*, pages 103–114. World Scientific.
- Albert, A. and Zhang, L. (2010). A novel definition of the multivariate coefficient of variation. *Biometrical Journal*, 52(5):667–675.

- Aldosari, M. S., Aslam, M., Khan, N., Ahmad, L., and Jun, C. H. (2018). A New S2Control Chart Using Multiple Dependent State Repetitive Sampling. *IEEE Access*, 6:49224–49236.
- Alkahtani, S. and Schaffer, J. (2012). A Double Multivariate Exponentially Weighted Moving Average (dMEWMA) Control Chart for a Process Location Monitoring. *Communications in Statistics - Simulation and Computation*, 41(2):238–252.
- Alodat, M. T., Al-Rawwash, M. Y., Nawajah, I. M., and Al-rawwash, M. Y. (2010). Inference about the Regression Parameters Using Median-Ranked Set Sampling. *Communications in Statistics—Theory and Methods*, 39:2604–2616.
- Aly, A. A., Mahmoud, M. A., and Hamed, R. (2016). The Performance of the Multivariate Adaptive Exponentially Weighted Moving Average Control Chart with Estimated Parameters. *Quality and Reliability Engineering International*, 32(3):957–967.
- Aly, A. A., Saleh, N. A., Mahmoud, M. A., and Woodall, W. H. (2015). A reevaluation of the adaptive exponentially weighted moving average control chart when parameters are estimated. *Quality and Reliability Engineering International*, 31(8):1611–1622.
- and others Ahmad, Shabbir and Lin, Zhengyan and Abbasi, Saddam Akber and Riaz, M. (2012). On Efficient Monitoring of Process Dispersion using Interquartile Range. *Open Journal of Applied Sciences*, 2(48):39–43.
- Anderson, M. J. (2001). Permutation tests for univariate or multivariate analysis of variance and regression. *Canadian Journal of Fisheries and Aquatic Sciences*, 58(3):626–639.
- Anderson, M. J., de Valpine, P., Punnett, A., and Miller, A. E. (2019). A pathway for multivariate analysis of ecological communities using copulas. *Ecology and Evolution*, (January):1–19.

- Anderson, M. J. and Ter Braak, C. J. (2003). Permutation tests for multi-factorial analysis of variance. *Journal of Statistical Computation and Simulation*, 73(2):85–113.
- Anderson, M. J. and Thompson, A. A. (2004). Multivariate control charts for ecological and environmental monitoring. *Ecological Applications*, 14(6):1921–1935.
- Anderson, M. J. and Willis, T. J. (2003). Canonical Analysis of Principal Coordinates: A Useful Method of Constrained Ordination for Ecology. *Source: Ecology Ecology*, 84(842):511–525.
- Ardia, David and Boudt, Kris and Gagnon Fleury, J.-P. (2017). RiskPortfolios : Computation of Risk-Based Portfolios in R. *The Journal of Open Source software*, 2(10):171.
- Aslam, M. (2018). Statistical Monitoring of Process Capability Index Having One Sided Specification under Repetitive Sampling Using an Exact Distribution. *IEEE Access*, 6:25270–25276.
- Aslam, M., Bantan, R. A., and Khan, N. (2019). Design of a control chart for gamma distributed variables under the indeterminate environment. *IEEE Access*, 7:8858–8864.
- Aslam, M., Bhattacharya, R., and Aldosari, M. S. (2018). Design of Control Chart in Presence of Hybrid Censoring Scheme. *IEEE Access*, 6:14895–14907.
- Aslam, M., Khan, N., and Jun, C. H. (2015). A new S2 control chart using repetitive sampling. *Journal of Applied Statistics*, 42(11):2485–2496.
- Babamoradi, H., Van Den Berg, F., and Rinnan, Å. (2016). Confidence limits for contribution plots in multivariate statistical process control using bootstrap estimates.
- Bajorski, P. (2011). *Statistics for Imaging, Optics, and Photonics*, volume 808.
- Balakrishnan, N., Triantafyllou, I. S., and Koutras, M. V. (2010). A Distribution-Free Control Chart Based on Order Statistics. *Communications in Statistics - Theory and Methods*, 39(20):3652–3677.

- Benneyan, J. C. (1998a). Statistical quality control methods in infection control and hospital epidemiology, Part II: Chart use, statistical properties, and research issues. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America*, 19(March):265–283.
- Benneyan, J. C. (1998b). Statistical quality control methods in infection control and hospital epidemiology, Part II: Chart use, statistical properties, and research issues. *Infection control and hospital epidemiology : the official journal of the Society of Hospital Epidemiologists of America*, 19(December):265–283.
- Bersimis, S., Psarakis, S., and Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, 23(5):517–543.
- Borrer, C. M., Montgomery, D. C., and Runger, G. C. (2018). Robustness of the EWMA Control Chart to Non-Normality. *Journal of Quality Technology*, 31(3):309–316.
- Brook, D. and Evans, D. A. (1972). An approach to the probability distribution of cusum run length. *Biometrika*, 59(1):539–549.
- Bundy, A. (2005). Structure and functioning of the eastern Scotian Shelf ecosystem before and after the collapse of groundfish stocks in the early 1990s. *Canadian Journal of Fisheries and Aquatic Sciences*, 62(7):1453–1473.
- Calzada, M. E. and Scariano, S. M. (2013). A synthetic control chart for the coefficient of variation. *Journal of Statistical Computation and Simulation*, 83(5):853–867.
- Castagliola, P., Achouri, A., Taleb, H., Celano, G., and Psarakis, S. (2013). Monitoring the coefficient of variation using control charts with run rules. *Quality Technology and Quantitative Management*, 10(1):75–94.
- Chakraborti, S., Van, P., Laan, D., and Bakir, S. T. (2001). Nonparametric control charts: An overview and some results. *Journal of Quality Technology*, 33(3).

- Champ, C. W. and Jones-Farmer, L. A. (2007). Properties of multivariate control charts with estimated parameters. *Sequential Analysis*, 26(2):153–169.
- Champ, C. W., Jones-Farmer, L. A., and Rigdon, S. E. (2005). Properties of the T2 control chart when parameters are estimated. *Technometrics*, 47(4):437–445.
- Chan, Lai K and Zhang, J. (2001). Cumulative sum control charts for the covariance matrix. *Statistica Sinica*, pages 767—790.
- Chen, Z. and Shen, L. (2003). Two-layer ranked set sampling with concomitant variables. *Journal of Statistical Planning and Inference*, 115:45–57.
- Cheng, Smiley W. and Thaga, K. (2006). Single Variables Control Charts: an Overview. *Quality and Reliability Engineering International*, 22(7):811—820.
- Choi, J. S., Frank, K. T., Leggett, W. C., and Drinkwater, K. (2004). Transition to an alternate state in a continental shelf ecosystem. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(4):505–510.
- Cochran, W. (1977). *Sampling Techniques*:. New York: Wiley, 3rd edition.
- Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30(3):291–303.
- Cui, P., Li, J., and Wang, G. (2008). Improved kernel principal component analysis for fault detection. *Expert Systems with Applications*, 34(2):1210–1219.
- David, H and Nagaraja, H. (2003). *Order Statistics 3rd ed.* John Wiley & Sons, London.
- De Vries, A. and Conlin, B. J. (2003). Design and performance of statistical process control charts applied to estrous detection efficiency. *Journal of dairy science*, 86(6):1970–84.

- De Vries, A. and Reneau, J. K. (2010). Application of statistical process control charts to monitor changes in animal production systems. *Journal of animal science*, 88(13 Suppl):11–24.
- Dell, T. R. and Clutter, J. L. (1972). Ranked set sampling theory with order statistics background. *Biometrics*, pages 545—555.
- Du, J. and Maceachern, S. N. (2008). Judgement Post-Stratification for Designed Experiments. *Source: Biometrics*, 64(2):345–354.
- Duncan, A. J. (1974). *Quality Control and Industrial Statistics*. 4th edition.
- El-Din, M. A. S., Rashed, H. I., and El-Khabeery, M. M. (2006). Statistical Process Control Charts Applied to Steelmaking Quality Improvement. *Quality Technology & Quantitative Management*, 3(4):473–491.
- Ellingsen, K. E., Anderson, M. J., Shackell, N. L., Tveraa, T., Yoccoz, N. G., and Frank, K. T. (2015). The role of a dominant predator in shaping biodiversity over space and time in a marine ecosystem. *Journal of Animal Ecology*, 84(5):1242–1252.
- Ewan, W. D. and Kemp, K. W. (1960). Sampling inspection of continuous processes with no autocorrelation between successive results. *Biometrika*, 47(7):363–380.
- Ferrer, A. (2007). Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process. *Quality Engineering*, 19(4):311—325.
- Ferrer, A. (2014). Latent Structures-Based Multivariate Statistical Process Control: A Paradigm Shift. *Quality Engineering*, 26(1):72—91.

- Frank, K. T., Petrie, B., Choi, J. S., and Leggett 2005, W. C. D. A. J. . (2005). Trophic Cascades in a Formerly Cod-Dominated Ecosystem. *Science*, 308(5728):1621–1623 ST – Trophic Cascades in a Formerly Cod.
- Frank, K. T., Petrie, B., Fisher, J. A. D., and Leggett, W. C. (2011). Transient dynamics of an altered large marine ecosystem. *Nature*, 477(7362):86–91.
- Frank, K. T., Petrie, B., and Shackell, N. L. (2007). The ups and downs of trophic control in continental shelf ecosystems. *Trends in Ecology and Evolution*, 22(5):236–242.
- Frank, K. T., Petrie, B., Shackell, N. L., and Choi, J. S. (2006). Reconciling differences in trophic control in mid-latitude marine ecosystems. *Ecology Letters*, 9(10):1096–1105.
- Franklin, S. B., Gibson, D. J., Robertson, P. A., Pohlmann, J. T., and Fralish, J. S. (1995). Parallel Analysis: A Method for Determining Significant Principal Components. *Source Journal of Vegetation Science Journal of Vegetation Science*, 6(6):99–106.
- Friedman, J. (1989). Regularized Discriminant Analysis. *Journal of the American Statistical Association*, 84(405):165–175.
- Frontier, S. (1976). Étude de la décroissance des valeurs propres dans une analyse en composantes principales: Comparaison avec le moddle du bâton brisé. *Journal of Experimental Marine Biology and Ecology*, 25(1):67–75.
- George Box (1988). Signal-to-Noise Ratios, Performance Criteria, and Transformations. *Technometrics*, 30(1):1–17.
- Giovanelli, J.-f. and Idier, J. (2015). *Regularization and Bayesian Methods for Inverse Problems in Signal and Image Processing*. Wiley Online Library.

- Golosnoy, V., Ragulin, S., and Schmid, W. (2009). Multivariate CUSUM chart: properties and enhancements. *AStA Adv Stat Anal*, 93:263–279.
- Gower, J. C. (1966). Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53(3-4):325–338.
- Greenland, S. (2000). Principles of multilevel modelling. *International Journal of Epidemiology*, 29(1):158–167.
- Grigg, Nigel P and Daly, Jeannette and Stewart, M. (1998). Case study: the use of statistical process control in fish product packaging. *Food control*, 9(5):289—298.
- Gupta, Arjun K and Nagar, D. K. (2018). Matrix Variate Distributions.
- Halls, Lowell K and Dell, T. R. (1966). Yields, Trial of ranked-set sampling for forage. *Forest Science*, 12(1):22—26.
- Haq, A., Brown, J., and Moltchanova, E. (2016). A New Synthetic Exponentially Weighted Moving Average Control Chart for Monitoring Process Dispersion. *Quality and Reliability Engineering International*, 32(1):241–256.
- Haq, A., Brown, J., Moltchanova, E., and Al-Omari, A. I. (2013). Partial ranked set sampling design. *Environmetrics*, 24(3):201–207.
- Hawkins, D. M. (1991). Multivariate Quality Control Based on Regression-Adjusted Variables. *Source: Technometrics on Thu TECHNOMETRICS*, 331239663(1):61–75.
- Hawkins, Douglas M and Maboudou-Tchao, E. M. (2007). Self-Starting Multivariate Exponentially Weighted Moving Average Control Charting. *Technometrics*, 49(2):199–209.

- Hayes, G. D., Scallan, A. J., and Wong, J. H. (1997). Applying statistical process control to monitor and evaluate the hazard analysis critical control point hygiene data. *Food Control*, 8(4):173–176.
- Hessainia, S., Bouchareb, F., Cheloufi, H., Berredjem, M., Berredjem, H., Becheker, I., and Aouf, N. E. (2013). Efficient synthesis and antibacterial activity of novel cyclic sulfamides. *Rasayan Journal of Chemistry*, 6(3):175–182.
- Higham, N. J. (1988). Computing a nearest symmetric positive semidefinite matrix. *Linear Algebra and Its Applications*, 103(C):103–118.
- Holmes, D. S. and Mergen, A. E. (1993). Improving the performance of the T2 control chart. *Quality Engineering*, 5(4):619—625.
- Hotelling, H. (1947). Multivariate quality control—illustrated by the air testing of sample bombsights. In *Techniques of Statistical Analysis (Ed MWHC Eisenhart, and WA Wallis. New York, NY, USA: Mc-Graw-Hill*, pages 111–184.
- Human, S. W., Kritzinger, P., and Chakraborti, S. (2011). Robustness of the EWMA control chart for individual observations. *Journal of Applied Statistics*, 38(10):2071–2087.
- Jackson, D. A. (1993). Stopping Rules in Principal Components Analysis : A Comparison of Heuristical and Statistical Approaches. *Source: Ecology*, 74(8):2204–2214.
- Jackson, J. E. (1991). *A User ' s Guide to Principal Components*. John Wiley & Sons, Inc.
- Jensen, W. A., Allison Jones-Farmer, L., Champ, C. W., and Woodall, W. H. (2006). Effects of Parameter Estimation on Control Chart Properties: A Literature Review. *Journal of Quality Technology*, 38(4).
- Johnson, N. L. (1961). A simple theoretical approach to cumulative sum control charts. *Journal of the American Statistical Association*, 56(296):835–840.

- Jones, L. A., Champ, C. W., and Rigdon, S. E. (2001). The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics*, 43(2):156—167.
- Kang, C. W., Lee, M. A. N. S., and Hawkins, D. M. (2007). A Control Chart for the Coefficient of Variation. *Journal of Quality Technology*, 39(2):151–158.
- Khaliq, Q. U. A., Riaz, M., and Ahmad, S. (2016). On designing a new Tukey-EWMA control chart for process monitoring. *International Journal of Advanced Manufacturing Technology*, 82(1-4):1–23.
- Khan, N., Aslam, M., Ahmad, L., and Jun, C. H. (2017). A control chart for gamma distributed variables using repetitive sampling scheme. *Pakistan Journal of Statistics and Operation Research*, 13(1):47–61.
- Khan, N., Aslam, M., Aldosari, M. S., and Jun, C. H. (2018). A multivariate control chart for monitoring several exponential quality characteristics using EWMA. *IEEE Access*, 6:70349–70358.
- Kim, J. (2014). *Change Point Detection in Univariate and Multivariate Processes*. Rutgers The State University of New Jersey New Brunswick.
- Knoth, S. (2017). ARL numerics for MEWMA charts. *Journal of Quality Technology*, 49(1):78–89.
- Knoth, S. (2018). Statistical Process Control – Calculation of ARL and Other Control Chart Performance Measures.
- Kourti, T. (2005). Application of latent variable methods to process control and multivariate statistical process control in industry. *International Journal of Adaptive Control and Signal Processing*, 19(4):213–246.
- Kourti, T. and Macgregor, J. F. (1996). Multivariate SPC Methods for Process and Product Monitoring. *Journal of quality technology*, 28(4):409—428.

- Kramer, H. G. and Schmid, L. (1997). Ewma charts for multivariate time series. *Sequential Analysis*, 16(2):131–154.
- Kruskal, J. B. (1964a). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27.
- Kruskal, J. B. (1964b). Nonmetric Multidimensional Scling: A Numerical Method. *PSYCHOMETRIKA*, 29(2).
- Lancewicki, T. and Aladjem, M. (2014). Multi-target shrinkage estimation for covariance matrices. *IEEE Transactions on Signal Processing*, 62(24):6380–6390.
- Ledoit, O. and Wolf, M. (2003). Improved estimation of the covariance matrix of stock returns with an application to portfolio selection. *Journal of empirical finance*, 10(5):603–621.
- Ledoit, O. and Wolf, M. (2004a). A well-conditioned estimator for large-dimensional covariance matrices. *Journal of Multivariate Analysis*, 88(2):365–411.
- Ledoit, O. and Wolf, M. (2004b). Honey, I Shrunk the Sample Covariance Matrix. *The Journal of Portfolio Management*, 30(4):110—119.
- Legendre, P. and Legendre, L. (1988). Numerical Ecology, Volume 24. (*Developments in Environmental Modelling*), 24:870.
- Lim, A. J., Khoo, M. B., Teoh, W. L., and Haq, A. (2017). Run sum chart for monitoring multivariate coefficient of variation. *Computers and Industrial Engineering*, 109:84–95.
- Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1):46–53.

- Lucas, J. M. and Crosier, R. B. (1982). Fast initial response for CUSUM quality-control schemes: give your CUSUM a head start. *Technometrics*, 24(3):199–205.
- Lucas, J. M. and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32(1):1–12.
- MacGregor, J. F., Jaeckle, C., Kiparissides, C., and Koutoudi, M. (1994). Process monitoring and diagnosis by multiblock PLS methods. *AIChE Journal*, 40(5):826–838.
- Madsen, T. N. and Kristensen, A. R. (2005). A model for monitoring the condition of young pigs by their drinking behaviour. *Computers and Electronics in Agriculture*, 48(2):138–154.
- Mahmoud, M. A. and Maravelakis, P. E. (2013). The performance of multivariate CUSUM control charts with estimated parameters. *Journal of Statistical Computation and Simulation*, 83(4):37–41.
- Mahmouda, M. A. and Maravelakisb, P. E. (2010). The performance of the MEWMA Control Chart When Parameters Are Estimated. *Communications in Statistics - Simulation and Computation*, 39(9):1803–1817.
- Mandel, B. J. (1969). The Regression Control Chart. *Journal of Quality Technology*, 1(1):1–9.
- Maravelakis, P., Panaretos, J., and Psarakis, S. (2004). EWMA Chart and Measurement Error. *Journal of Applied Statistics*, 31(4):445–455.
- Maravelakis, P., Panaretos, J., and Psarakis, S. (2005). An examination of the robustness to non normality of the EWMA control charts for the dispersion. *Communications in Statistics: Simulation and Computation*, 34(4):1069–1079.
- Maravelakis, P. E., Bersimis, S., Panaretos, J., and Psarakis, S. (2002). Identifying the out of con-

- trol variable in a multivariate control chart. *Communications in Statistics - Theory and Methods*, 31(12):2391–2408.
- Marlin, T. E. (2000). *Process Control: Designing Processes and Control Systems for Dynamic Performance*.
- Martin, E. B. and Morris, A. J. (1996). Non-parametric confidence bounds for process performance monitoring charts. *Journal of Process Control*, 6(6):349–358.
- May, G. S. and Spanos, C. J. (2006). *Fundamentals of semiconductor manufacturing and process control*. John Wiley & Sons.
- McCune, Bruce and Grace, James B and Urban, D. L. (2002). *Analysis of ecological communities*. MjM software design Gleneden Beach, OR.
- McIntyre, G. A. (1952). A Method for Unbiased Selective Sampling, Using Ranked Sets. *Australian Journal of Agricultural Research*, 3(4):385—390.
- Mehmood, R., Riaz, M., and Does, R. J. M. M. (2013). Control charts for location based on different sampling schemes. *Journal of Applied Statistics*, 40(3):483–494.
- Melrose, J., Perroy, R., and Careas, S. (2015). *High-Dimensional Covariance Estimation: With High-Dimensional Data*, volume 1.
- Miller, L. D., Smeds, J., George, J., Vega, V. B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E. T., and Bergh, J. (2005). From The Cover: An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proceedings of the National Academy of Sciences*, 102(38):13550–13555.
- Montgomery, D. C. (2009). *Statistical Quality Control: A Modern Introduction*.

- Montgomery, D. C. and Woodall, W. H. (1997). A discussion on statistically-based process monitoring and control. *Journal of Quality Technology*, 29(2):121.
- Morris, C. N. (1983). Parametric Empirical Bayes Inference: Theory and Applications. *Journal of the American Statistical Association*, 78(381):47–55.
- Murthy, M. N. (1964). Product method of estimation. *The Indian Journal of Statistics, Series A*, 26(1):69—74.
- Muttlak, H. (1997). Median ranked set sampling. *Journal of Applied Statistical Sciences*, 6(4):245—255.
- Ngai, H.-M. and Zhang, J. (2001). Multivariate cumulative sum control charts based on projection pursuit. *Statistica Sinica*, 11:747–766.
- Opgein, R. and Strimmer, K. (2007). Statistical Applications in Genetics and Molecular Biology. *Statistical Applications in Genetics and ...*, 6(1).
- Ou, Y., Chen, N., and Khoo, M. B. (2015). An efficient multivariate control charting mechanism based on SPRT. *International Journal of Production Research*, 53(7):1937–1949.
- Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41(1/2):100–115.
- Page, E. S. (1961). Cumulative sum charts. *Technometrics*, 3(1):1–9.
- Park, J. and Jun, C. H. (2015). A new multivariate EWMA control chart via multiple testing. *Journal of Process Control*, 26:51–55.
- Patil, G. P., El-Shaarawi, A. H., and Piegorsch, W. W. (2002). Ranked set sampling. *Encyclopedia of environmetrics*, 3:1684–1690.
- Patil, GP and Sinha, AK and Taille, C. (1993). Relative precision of ranked set sampling: a comparison with the regression estimator. *Environmetrics*, 4(4):399—412.

- Paul, W. L. and Anderson, M. J. (2013). Causal modeling with multivariate species data. *Journal of Experimental Marine Biology and Ecology*, 448:72–84.
- Pawitan, Y., Bjöhle, J., Amler, L., Borg, A. L., Egyhazi, S., Hall, P., Han, X., Holmberg, L., Huang, F., Klaar, S., Liu, E. T., Miller, L., Nordgren, H., Ploner, A., Sandelin, K., Shaw, P. M., Smeds, J., Skoog, L., Wedrén, S., and Bergh, J. (2005). Gene expression profiling spares early breast cancer patients from adjuvant therapy: derived and validated in two population-based cohorts. *Breast cancer research : BCR*, 7(6):R953–R964.
- Pereira, M. A., Weggemans, R. M., Jacobs, D. R., Hannan, P. J., Zock, P. L., Ordovas, J. M., and Katan, M. B. (2004). Within-person variation in serum lipids: implications for clinical trials. *International Journal of Epidemiology*, 33(3):534–541.
- Pettersson, M. (1998). Monitoring a freshwater fish population: Statistical surveillance of biodiversity. *Environmetrics*, 9(2):139–150.
- Phaladiganon, P., Kim, S. B., Chen, V. C. P., and Jiang, W. (2013). Principal component analysis-based control charts for multivariate nonnormal distributions. *Expert Systems With Applications*, 40:3044–3054.
- Pignatiello, J. J. and Runger, G. C. (1990). Comparisons of multivariate CUSUM charts. *Journal of Quality Technology*, 22(3):173–186.
- Pongpullponsak, A. and Sontisamran, P. (2013). Statistical Quality Control Based on Ranked Set Sampling for Multiple Characteristics. *Chiang Mai J. Sci.*, 40(403):485–498.
- Qiu, P. (2008). Distribution-free multivariate process control based on log-linear modeling. *IIE Transactions*, 40:664–677.

- Qiu, P. and Hawkins, D. (2001). A Rank-Based Multivariate CUSUM Procedure. *Technometrics*, 43(2):120–132.
- Qiu, P. and Hawkins, D. (2003). A Nonparametric Multivariate Cumulative Sum Procedure for Detecting Shifts in All Directions. *Journal of the Royal Statistical Society. Series D (The Statistician)*, 52(2):151–164.
- Quesenberry, C. P. (1991). SPC Q charts for start-up processes and short or long runs. *Journal of quality technology*, 23(3):213–224.
- Quesenberry, C. P. (1997). *SPC methods for quality improvement*.
- R Core Team (2013). R: A language and environment for statistical computing.
- Reed, G. F., Lynn, F., and Meade, B. D. (2002). Use of Coefficient of Variation in Assessing Variability of Quantitative Assays. *Clinical and diagnostic laboratory immunology*, 9(6):1235–1239.
- Reyment, R. A. (1960). *Studies on Nigerian Upper Cretaceous and Lower Tertiary Ostracoda. P. 1, Senonian and Maestrichtian Ostracoda*. Almqvist & Wiksell.
- Riaz, M. (2008a). Monitoring process mean level using auxiliary information. *Statistica Neerlandica*, 62(4):458–481.
- Riaz, M. (2008b). Monitoring process variability using auxiliary information. *Computational statistics*, 23(2):253–276.
- Riaz, M. (2015). On enhanced interquartile range charting for process dispersion. *Quality and Reliability Engineering International*, 31(3):389–398.
- Riaz, M., Abbas, N., and Does, R. J. M. M. (2011). Improving the performance of CUSUM charts. *Quality and Reliability Engineering International*, 27(4):415–424.

- Riaz, M., Abbasi, S. A., Ahmad, S., and Zaman, B. (2014). On efficient phase II process monitoring charts. *International Journal of Advanced Manufacturing Technology*, 70(9-12):2263–2274.
- Riaz, M., Mehmood, R., Ahmad, S., and Abbasi, S. A. (2013). On the Performance of Auxiliary-based Control Charting under Normality and Nonnormality with Estimation Effects. *Quality and Reliability Engineering International*, 29(8):1165—1179.
- Riaz, M. and Schoonhoven, M. (2011). Design and Analysis of Control Charts for Standard Deviation with Estimated Parameters. *Journal of Quality Technology*, 43(4):307—333.
- Ridwan A. Sanusi, N. A. & M. R. (2017). On efficient CUSUM-type location control charts using auxiliary information. *Quality Technology & Quantitative Management*, 0(0):1—19.
- Roberts, S. W. (1959). Control chart tests based on geometric moving averages. *Technometrics*, 1(3):239–250.
- Robson, D. S. (1957). Applications of multivariate polykeys to the theory of unbiased ratio-type estimation. *Journal of the American Statistical Association*, 52(280):511–522.
- Ronning, G. (1977). A Simple Scheme for Generating Multivariate Gamma Distributions with Non-Negative Covariance Matrix. *American Society for Quality*, 19(2):179–183.
- Saleh, Nesma A and Mahmoud, M. A. (2017). Accounting for phase I sampling variability in the performance of the MEWMA control chart with estimated parameters. *Communications in Statistics-Simulation and Computation*, 46(6):1—15.
- Samawi, H. M. and Tawalbeh, E. M. (2002). Double Median Ranked Set Sample: Comparing To Other Double Ranked Samples For Mean And Ratio Estimators. *Journal of Modern Applied Statistical Methods*, 1(2):428–442.

- Samawi, Hani M and Ahmed, Mohammad S and Abu-Dayyeh, W. (1996). Estimating the Population Mean Using Extreme Ranked Set Sampling. *Biometrical Journal*, 38(5):577—586.
- Samawi, Hani M and Muttlak, H. A. (1996). Estimation of Ratio Using Rank Set Sampling. *Biometrical Journal*, 38(6):753–764.
- Santos-Fernandez, E. (2012). *Multivariate Statistical Quality Control Using R*. Springer Science & Business Media, New York.
- Santos-Fernandez, Edgar and Santos-Fernandez, M. E. (2016). *Package ‘MSQC’*.
- Sanusi, R. A., Abujiya, M. R., Riaz, M., and Abbas, N. (2017a). Combined Shewhart CUSUM charts using auxiliary variable. *Computers & Industrial Engineering*, 105:329–337.
- Sanusi, R. A., Riaz, M., Abbas, N., and Abujiya, M. R. (2016). Using FIR to Improve CUSUM Charts for Monitoring Process Dispersion. *Quality and Reliability Engineering International*, 33(5):1045—1056.
- Sanusi, R. A., Riaz, M., Adegoke, N. A., and Xie, M. (2017b). An EWMA monitoring scheme with a single auxiliary variable for industrial processes. *Computers & Industrial Engineering*, 114:1–10.
- Schäfer, J. and Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical applications in genetics and molecular biology*, 4(1).
- Schipper, M., Den Hartog, J., and Meelis, E. (1997). Sequential analysis of environmental monitoring data: optimal SPRTs. *Environmetrics*, 8(1):29—41.
- Seif, A., Moghadam, M. B., Faraz, A., and Heuchenne, C. (2011). Statistical Merits and Economic Evaluation of T2 Control Charts with the VSSC Scheme. *Arabian Journal for Science and Engineering*, 36(7):1461–1470.

- Shabbir, J. and Awan, W. H. (2016). An Efficient Shewhart-Type Control Chart to Monitor Moderate Size Shifts in the Process Mean in Phase II. *Quality and Reliability Engineering International*, 32(5):1597–1619.
- Shackell, N. L. and Frank, K. T. (2003). Marine fish diversity on the Scotian Shelf, Canada. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 13(4):305–321.
- Shackell, N. L., Frank, K. T., Fisher, J. A., Petrie, B., and Leggett, W. C. (2010). Decline in top predator body size and changing climate alter trophic structure in an oceanic ecosystem. *Proceedings of the Royal Society B: Biological Sciences*, 277(1686):1353–1360.
- Shewhart, W. A. (1931). *Economic control of quality of manufactured product*, volume 509. ASQ Quality Press.
- Shi, X., Lv, Y., Fei, Z., and Liang, J. (2013). A multivariable statistical process monitoring method based on multiscale analysis and principal curves. *International Journal of Innovative Computing*, 9(4):1781–1800.
- Siegmund, D. (1985). *Sequential analysis: tests and confidence intervals*. Springer, New York.
- Singh, Ravindra and Mangat, N. S. (2013). *Elements of survey sampling*. Springer Science & Business Media, Dordrecht.
- Srikaeo, K. and Hourigan, J. A. (2002). The use of statistical process control (SPC) to enhance the validation of critical control points (CCPs) in shell egg washing. *Food Control*, 13(4-5):263–273.
- Stein, C. and Others (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution. In *Proceedings of the Third Berkeley symposium on mathematical statistics and probability*, volume 1, pages 197–206.

- Steneck, R. S. (2012). Apex predators and trophic cascades in large marine ecosystems: Learning from serendipity. *Proceedings of the National Academy of Sciences*, 109(21):7953–7954.
- Stokes, S. L. (1977). Ranked set sampling with concomitant variables. *Communications in Statistics-Theory and Methods*, 6(12):1207—1211.
- Stoumbos, Z. G. and Reynolds, M. R. (2000). Robustness of non-normality and autocorrelation of individuals control charts. *Journal of Statistical Computation and Simulation*, 66(2):145–187.
- Stoumbos, Z. G. and Sullivan, J. H. (2002). Robustness to Non-normality of the Multivariate EWMA Control Chart. *Journal of Quality Technology*, 34(3):260–276.
- Strang, G. (2009). *Introduction to Linear Algebra, Fourth Edition*.
- Sullivan, J. H. and Jones, L. A. (2002). A Self-Starting Control Chart for Multivariate Individual Observations. *Technometrics*, 44(1):24—33.
- Sullivan, J. H. and Woodall, W. H. (1996). A comparison of multivariate control charts for individual observations. *Journal of Quality Technology*, 28(4):398–408.
- Takahasi, K. and Wakimoto, K. (1968). On unbiased estimates of the population mean based on the sample stratified by means of ordering. *Annals of the Institute of Statistical Mathematics*, 20(1):1–31.
- Teoh, Wei Lin and Khoo, Michael BC and Castagliola, Philippe and Lee, M. H. (2016). The Exact Run Length Distribution and Design of the Shewhart X Chart with Estimated Parameters Based on Median Run Length. *Communications in Statistics—Simulation and Computation*, 45:2081–2103.
- Terlizzi, A., Benedetti-Cecchi, L., Bevilacqua, S., Fraschetti, S., Guldetti, P., and Anderson, M. J. (2005a). Multivariate and univariate asymmetrical analyses in environmental impact assessment: A case study of Mediterranean subtidal sessile assemblages. *Marine Ecology Progress Series*, 289(May 2014):27–42.

- Terlizzi, A., Scuderi, D., Fraschetti, S., and Anderson, M. J. (2005b). Quantifying effects of pollution on biodiversity: A case study of highly diverse molluscan assemblages in the Mediterranean. *Marine Biology*, 148(2):293–305.
- Testik, M. C., Runger, G. C., and Borror, C. M. (2003). Robustness properties of multivariate EWMA control charts. *Quality and Reliability Engineering International*, 19(1):31–38.
- Thor, A. D., Berry, D. A., Budman, D. R., Muss, H. B., Kute, T., Henderson, I. C., Barcos, M., Cirincione, C., Edgerton, S., Allred, C., Norton, L., and Liu, E. T. (1998). erbB-2, p53, and efficacy of adjuvant therapy in lymph node-positive breast cancer. *Journal of the National Cancer Institute*, 90(18):1346–1360.
- Tsung, C. Z. & F. and To (2012). A Multivariate Sign EWMA Control Chart. *Technometrics*, 53(1):84–97.
- Tuerhong, G. and Bum Kim, S. (2014). Gower distance-based multivariate control charts for a mixture of continuous and categorical variables. *Expert Systems With Applications*, 41:1701–1707.
- Ullah, I. (2015). *Contributions to high-dimensional data analysis: some applications of the regularized covariance matrices*. PhD thesis, Massey University.
- Ullah, I., Pawley, M., Smith, A., and Jones, B. (2017). Improving the detection of unusual observations in high-dimensional settings. *Australian and New Zealand Journal of Statistics*, 59(4):449–462.
- Van Valen, L. (1974). Multivariate structural statistics in natural history. *Journal of Theoretical Biology*, 45(1):235–247.
- Voinov, V. G. and Nikulin, M. S. (1996). *Unbiased Estimators and Their Applications: Multivariate Case.*, volume 2. Kluwer.

- Wang, T. and Huang, S. (2016). An adaptive multivariate CUSUM control chart for signaling a range of location shifts. *Communications in Statistics - Theory and Methods*, 45(16):4673–4691.
- Warton, D. I. (2008). Penalized normal likelihood and ridge regularization of correlation and covariance matrices. *Journal of the American Statistical Association*, 103(481):340–349.
- Whittaker, J. (2009). *Graphical models in applied multivariate statistics*. Wiley Publishing.
- Woodall, W. H. (1983). The distribution of the run length of one-sided CUSUM procedures for continuous random variables. *Technometrics*, 25(3):295–301.
- Woodall, W. H. and Mahmoud, M. A. (2005). The inertial properties of quality control charts. *Technometrics*, 47(4):425–436.
- Woodall, W. H. and Ncube, M. M. (1985). Multivariate CUSUM Quality- Control Procedures. *Source: Technometrics UTC TECHNOMETRICS*, 27(3):285–29242.
- Worm, B. and Myers, R. A. (2003). Meta-Analysis of Cod-Shrimp Interactions Reveals Top-Down Control in Oceanic Food Webs Author (s): Boris Worm and Ransom A . Myers Published by : Wiley Stable URL : <http://www.jstor.org/stable/3108006> Accessed : 22-08-2016 10 : 26 UTC Your use of the JS. *Ecology Society of America*, 84(1):162–173.
- Wu, Z., Jiao, J., Yang, M., Liu, Y., and Wang, Z. (2009). An enhanced adaptive CUSUM control chart. *IIE Transactions*, 41(7):642–653.
- Yen, C. L. and Shiau, J. J. H. (2010). A Multivariate Control Chart for Detecting Increases in Process Dispersion. *Statistica Sinica*, 20(4):1683–1707.
- Yeong, W., Khoo, M., Tham, L., Teoh, W., and Rahim, M. (2017). Monitoring the coefficient of variation using a variable sampling interval ewma chart. *Journal of Quality Technology*, 49(4).

- Yeong, W. C., Khoo, M. B. C., Teoh, W. L., and Castagliola, P. (2016). A Control Chart for the Multivariate Coefficient of Variation. *Quality and Reliability Engineering International*, 32(3):1213–1225.
- Yoon, S. and Macgregor, J. F. (2001). Fault diagnosis with multivariate statistical models part I: using steady state fault signatures. *Journal of Process Control*, 11(4):87—400.
- Yumin, L. (1996). An Improvement for Mewma in Multivariate Process Control. *Computers and Industrial Engineering*, 31(3):779–781.
- Zhang, G. (1985). Cause-selecting control charts – a new type of quality control charts. *QR Journal*, 12(1):21–25.
- Zhang, G. and Chang, S. I. (2008). Multivariate EWMA control charts using individual observations for process mean and variance monitoring and diagnosis. *International Journal of Production Research*, 46(24):6855–6881.
- Zhang, J., Li, Z., Chen, B., and Wang, Z. (2014). A new exponentially weighted moving average control chart for monitoring the coefficient of variation. *Computers and Industrial Engineering*, 78:205–212.
- Zhang, X., Liu, L., Li, M., Chang, Y., Shang, L., Dong, J., Xiao, L., and Ao, Y. (2016). Improving the interfacial properties of carbon fibers/vinyl ester composites by vinyl functionalization on the carbon fiber surface. *RSC Adv.*, 6(35):29428–29436.
- Zhou, C., Zou, C., Zhang, Y., and Wang, Z. (2007). Nonparametric control chart based on change-point model. *Stat Papers*, 50:13–28.
- Zou, C., Wang, Z., and Tsung, F. (2012). A Spatial Rank-Based Multivariate EWMA Control Chart. *Naval Research Logistics (NRL)*, 59(2):91—110.

Zwanenburg, K. C. (2000). The effects of fishing on demersal fish communities of the Scotian Shelf. *ICES Journal of Marine Science*, 57(3):503–509.

Appendices

Appendix A

Other research involvements:

A.1 Co-author papers

.

- 1 Sanusi, R. A., Riaz, M., Adegoke, N. A., & Xie, M. (2017). An EWMA monitoring scheme with a single auxiliary variable for industrial processes. *Computers & Industrial Engineering*, 114, 1-10
- 2 Abbasi, S. A., Abbas, T., & Adegoke, N. A. (2019). Efficient CV Control Charts Based on Ranked Set Sampling. *IEEE Access*, 7, 78050-78062.

A.2 Conferences

- Nurudeen A. Adegoke, Marti J. Anderson, Adam N.H. Smith, Matthew D. M. Pawley. A change point distance-based multivariate control chart tool for ecological and environmental monitoring. JSM 2019, USA.
- Nurudeen A. Adegoke, Marti J. Anderson, Adam N.H. Smith, Matthew D. M. Pawley. Distance-

based multivariate control charts for ecological Monitoring, INMS Postgraduate Conference, New Zealand, 2018.

- Adegoke, N.A., Riaz, M., Sanusi, R.A., Smith, A.N.H., Pawley, M.D.M. EWMA-type Scheme for Monitoring Location Parameter using Auxiliary Information, INMS Postgraduate Conference, New Zealand, 2017.
- Nurudeen A. Adegoke, Mat Pawley, Adam Smith. Control Chart Schemes for High Dimensional Monitoring. Joint Conference of the New Zealand Statistical Association and the Operations Research Society of New Zealand. Nov 28 – 30 2016.
- Nurudeen A. Adegoke, Mat Pawley, Marti Anderson and Adam Smith. Control Charts Scheme for High Dimensional Monitoring. INMS Postgraduate Conference, New Zealand, 2016.

Appendix B

Formulas and Derivations

B.1 Derivation of the mean vector and covariance matrix of H_i

From equation (5.9), we have that for an in-control situation the mean vector of H_i is given as:

$$E(\mathbf{H}_i) = wE(\mathbf{Y}_i) + (1 - w)E(\bar{\mathbf{Y}}_{i-1})$$

$$E(\mathbf{H}_i) = w(\boldsymbol{\mu}_0) + (1 - w)(\boldsymbol{\mu}_0)$$

$$E(\mathbf{H}_i) = \boldsymbol{\mu}_0.$$

The covariance matrix of H_i is given as: when $i = 1$, we have

$$\mathbf{H}_1 = w\mathbf{y}_1 + (1 - w)\boldsymbol{\mu}_0$$

$$Var(\mathbf{H}_i) = w^2\boldsymbol{\Sigma}_0 + (1 - w)^2Var(\boldsymbol{\mu}_0)$$

$$Var(\mathbf{H}_i) = w^2 \mathbf{\Sigma}_0$$

when $i > 1$, we have:

$$Var(\mathbf{H}_i) = w^2 Var(\mathbf{Y}_i) + (1 - w)^2 Var(\bar{\mathbf{Y}}_{i-1}) + 2w(1 - w)Cov(\mathbf{Y}_i, \bar{\mathbf{Y}}_{i-1})$$

where, we have assumed that Y_i are independent and identical distributed. Hence, $Cov(\mathbf{Y}_i, \bar{\mathbf{Y}}_{i-1}) = 0$ for all pair of i and $i - 1$.

$$Var(\mathbf{H}_i) = w^2 \mathbf{\Sigma}_0 + (1 - w)^2 \frac{\mathbf{\Sigma}_0}{(i - 1)},$$

Hence, the covariance matrix of H_i is given as:

$$\Sigma_{H_i}^2 = \begin{cases} w^2 \mathbf{\Sigma}_0 & \text{if } i = 1 \\ w^2 \mathbf{\Sigma}_0 + (1 - w)^2 \frac{\mathbf{\Sigma}_0}{(i - 1)} & \text{if } i > 1 \end{cases} \quad (\text{B.1})$$

B.2 Proof of the non-centrality parameter

This proof that the distribution of the MHWMA test statistic H_i depends only on the value of the non-centrality parameter is based on the proof in Crosier (1988) and Lowry et al. (1992). The basic idea is to show that the values of H_i are invariant to any full-rank transformation of the data. That is, if \mathbf{M} is a $p \times p$ full rank matrix and $y^* = My$, then the MHWMA statistics \mathbf{H}_i , and also, the T^2 value, have the same value when calculated from y^* as when calculated from y . Hence, $\mathbf{H}^* = \mathbf{M}\mathbf{H}$. Crosier (1988) have chosen an orthogonal matrix \mathbf{M} that diagonalizes $\mathbf{\Sigma}_0$. From equation (5.9), when $i = 1$, we have

$$\mathbf{H}_1^* = \mathbf{M}(w\mathbf{y}_1 + (1 - w)\boldsymbol{\mu}_0) = \mathbf{M}\mathbf{H}_1$$

Hence, it follows that

$$T_1^{*2} = H_1^{*'} \Sigma_{H_1}^{-1} H_1^*$$

$$T_1^{*2} = H_1' M' (M'^{-1} \Sigma_{H_1}^{-1} M^{-1}) M H_1$$

where, $M' M'^{-1} = M^{-1} M = I$

$$T_1^{*2} = H_1' \Sigma_{H_1}^{-1} H_1 = T_1^2$$

When $i > 1$, we have:

$$H_i^* = M(w y_i + (1 - w) \bar{y}_{i-1}) = M H_i$$

Hence, it follows that

$$T_i^{*2} = H_i^{*'} \Sigma_{H_i}^{-1} H_i^*,$$

$$T_i^{*2} = H_i' M' (M'^{-1} \Sigma_{H_i}^{-1} M^{-1}) M H_i$$

$$T_i^{*2} = H_i' \Sigma_{H_i}^{-1} H_i = T_i^2$$

The results in Crosier (1988) can now be applied.

Appendix C

R Scripts

C.1 R code for the corrected limits in Chapter 3

```
library(spc)

library("mvtnorm")

library("MASS")

library("corpcor")

library("Matrix")

library(RiskPortfolios)

lmb=c(0)

mewma2<-function(sim,r,p,ucl,row,

cov_method=c("empir", "MSSD", "SW", "shrinkMedian"),

struct_mat =c("identity", "struct")){

  col=p ;  meansmodule <- rep(0,p)

  if(struct_mat=="identity"){

    V =diag(p)
```

```

}else{

  A <- matrix(0.8, p, p)

  V <- A^matrix(abs(col(A) - row(A)), p, p)

}

rlt=matrix(0,sim,1);arl=c();sdt=c()

for(b in 1:length(lmb)) {

  rl<-c()

  for(i in 1:sim){

    mmk=matrix(0, nrow=row, ncol=p)

    for(v in 1:row){

      mmk[v,]<- rmvnorm(1,mean=meansmodule,sigma=V)

    }

    if(cov_method=="empir"){

      covariance= cov(mmk)

    }else if(cov_method=="MSSD"){

      H <- matrix(0, row - 1, p)

      w1 = mmk

      for (ki in 1:row - 1) {

        H[ki, ] <- w1[ki + 1, ] - w1[ki, ]

      }

      covariance <- 0.5 * t(H) %*% H/(row - 1)

    }else if(cov_method=="SW"){

      sw <- matrix(0, p, p)

      w1 <- sweep(mmk, 2, (apply(mmk, 2, mean)))

```

```

for (ki in 1:row) {

  sw <- sw + w1[ki, ] %*% t(w1[ki, ])

}

covariance = sw/(row - 1)

}else {

  covariance= cov.shrink(mmk, verbose = "FALSE")

}

szi<- (r/(2-r))*covariance ; ii<-solve(szi)

zo<- colMeans(mmk); m1=sqrt(lmb[b]^2/p) ; mean = rep(m1,p)

vi<-c();x<-c(); xb<-c()

cnt<-0; k1<-0; k2<-1

x<-rmvnorm(1,mean=mean,sigma=V)

xbar = t(x)

z<-r*(xbar)+(1-r)*zo

vi1<- t(z-zo)%*%ii%*(z-zo)

if(vi1>ucl){cnt<-1}

while(cnt<1){

  x<-rmvnorm(1,mean=mean,sigma=V)

  xbar = t(x)

  z<-r*(xbar)+(1-r)*z

  vi<- t(z-zo)%*%ii%*(z-zo)

  if(vi>ucl){cnt<-1}

  k1<-k1+1

  k2<-k2+1

```

```

    }

    rl[i]<-k2

  }

  rlt[,b]=rl ; arl[b] = mean(rl);   sdt[b]=sd(rl)

}

return(arl)

}

```

```

BiSearch <- function(AARL, sim,r,p,row, cov_method=c("empir", "MSSD", "SW",
"shrinkMedian"), struct_mat =c("identity", "struct")) {

  L0=AARL;   c2 = mewma.crit(l=r, L0=AARL, p=p, hs=0, r=50)

  L2 = mewma2(sim,r, p, ucl= c2,row, cov_method, struct_mat);

  while(L2 < L0){

    c1 = c2; L1 = L2;    c2 =c2+ 1.;

    L2 = mewma2(sim,r, p, ucl= c2,row, cov_method, struct_mat);

    # print(round(c(c1,c2,L2),3))

  }

  c3 = c1 + (L0-L1)/(L2-L1) * (c2-c1);

  L3 = mewma2(sim,r, p, ucl= c3,row, cov_method, struct_mat);

  dc = c3 - c2;c1 = c2; L1 = L2; c2 = c3; L2 = L3;

  while((abs(L0-L3)> 0.1)){

    c3 = c1 + (L0-L1)/(L2-L1) * (c2-c1);

    L3 = mewma2(sim,r, p, ucl= c3,row, cov_method, struct_mat);

    dc = c3 - c2; c1 = c2; L1 = L2; c2 = c3; L2 = L3;
  }
}

```



```

    # print(round(c(c3,L0, L3),3))

}

xn=matrix(round(cbind(c3, L0, L3),4),1,3)

colnames(xn)=c("corrected limit", "ARL of the known parameter", "ARL of the corrected limit")

rownames(xn)=row

return(xn)

}

sim1=20000;

# The parameters below should be modified to appropriately.

AARL1=200; # this fix ARL to 200

r1=0.05; # the MEWMA chart smoothing parameter

p1=2; # number of variables

row1=30 # number of Phase I sample

p2_row_30=BiSearch(AARL=AARL1, sim=sim1,r=r1,p=p1,row=row1, cov_method="empir",

struct_mat ="identity")

```

C.2 R code of the results (including ARL and SDRL) of the MHWMA chart in Chapter 5

```

library(mvtnorm)

Norm=function(p, mu,sig0,w,UCL,sim){

  shift= c(0, 0.05, 0.1, 0.25,0.5, 0.75,1, 1.5,2,2.5,3,5)

  rlt=matrix(0,sim,length(shift));arl=c();sdt=c()

  mdn=c(); perc25=c(); perc75=c(); perc90=c()

```

```

for(b in 1:length(shift)) {

  m1=sqrt(shift[b]^2/p) ;      mu1=rep(m1,p)

  rlen = c()

  for(i in 1:sim){

    x=matrix(0, nrow=100000,ncol= p)

    ewma=matrix(0, nrow=100000,ncol= p)

    vi1=c()

    cnt = 0; rl = 1

    x[1,]= rmvnorm(1,mu1, sig0)

    ewma[1,]= w*x[1,]+ (1-w)*mu0

    szi<-w^2*sig0 ; ii<-solve(szi)

    vi1[1]<-sum( ewma[1,]*(ii%% ewma[1,]))

    if(vi1[1] > UCL) {cnt = 1}

    while(cnt < 1){

      rl = rl + 1

      x[rl,]= rmvnorm(1,mu1, sig0)

      summs=apply(x[1:rl,], 2, cumsum)

      ewma[rl,]= w*x[rl,]+ (1-w)* (summs[rl-1,]/(rl-1)) #colMeans(x[-i,]) #

      szi<-(w^2*sig0) + (1-w)^2*(sig0/(rl-1))

      ii<-solve(szi)

      vi1[rl]<-sum( ewma[rl,]*(ii%% ewma[rl,]))

      if(vi1[rl] > UCL) {cnt = 1}

    }

    rlen[i] = rl
  }
}

```

```

        #print(mean(rlen))

    }

    rlt[,b]= rlen

    arl[b] = mean(rlen)

    sdt[b]=sd(rlen)

    mdn[b]=median(rlen)

    perc25[b]=quantile(rlen,0.25)

    perc75[b]=quantile(rlen,0.75)

    perc90[b]=quantile(rlen,0.90)

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

yn= rlt

list(xn,yn)

}

p1=2; w1=0.03; UCL1=5.40

mu0 = rep(0,p1); sig0=diag(p1)

sim1=5000

results=Norm(p=p1,mu=mu0, sig0=sig0,w=w1, UCL=UCL1, sim =sim1)[[1]]

results

```

C.3 R-codes for the analysis in Chapter 9

C.3.1 R-codes Classical CUSUM, EWMA and HWMA charts

Classical CUSUM

```
library(MASS); library(spc)

CUSUM=function(K,H,sigy, muy, n,sim){

  shift= c(0, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.5, 0.75, 1, 1.5, 2)

  ybar= c() ; rlt=matrix(0,sim,length(shift));arl=c();sdt=c() ; mdn=c();

  perc25=c(); perc75=c(); perc90=c() ; mup=muy; sdp = sqrt((sigy^2/n))

  for(b in 1:length(shift)) {

    rlen=c()

    for(j in 1:sim)

    {

      m1= muy + shift[b]*(sigy/sqrt(n))

      c1 = c2 =c()

      for(i in 1:100000)

      {

        y = rnorm(n, m1, sigy); ybar = mean(y)

        ybar= (ybar-mup)/sdp

        if(i==1)

        {

          c1[i] = max(0,(ybar-0) - K + 0)

          c2[i] = max(0,-(ybar-0) - K + 0)

        }

      }

    }

  }

}
```

```

else

{

    c1[i] = max(0,(ybar-0) - K + c1[i-1])

    c2[i] = max(0,-(ybar-0) - K + c2[i-1])

}

if(c1[i] > H || c2[i]>H)

{

    rlen[j]=i

    #print(cbind(length(r1),r1[j],mean(r1)))

    break

}

else

{

    rlen[j]=100000

}

}

}

rlt[,b]= rlen

arl[b] = mean( rlen)

sdt[b]=sd( rlen)

mdn[b]=median( rlen)

perc25[b]=quantile( rlen,0.25)

perc75[b]=quantile( rlen,0.75)

perc90[b]=quantile( rlen,0.90)

```

```

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

yn= rlt

list(xn,yn)

}

sim1=50000; muy1=0; sigy1=1 ; mux1=0; sigx1=1 ;n1=1

K1=0.125; H1=xcusum.crit(k=K1, L0=500, mu0 = 0, hs = 0, sided = "two", r = 30)

K1_0_125=CUSUM(K=K1,H=H1,sigy=sigy1, muy=muy1, n=n1, sim=sim1)[[1]]

```

Classical EWMA chart

```

EWMA=function(muy,sigy ,lmb,L,sim, n){

shift= c(0, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.5, 0.75, 1, 1.5, 2)

rlt=matrix(0,sim,length(shift));arl=c();sdt=c() ; mdn=c(); perc25=c(); perc75=c(); perc90=c()

CL = UCL = LCL= double()

mu0=muy; s0 = sqrt((sigy^2/n)) ; t = 1:1e5

LCL[t] = mu0 - L * s0*sqrt(lmb*(1-(1-lmb)^(2*t))/(2-lmb))

UCL[t] = mu0 + L * s0*sqrt(lmb*(1-(1-lmb)^(2*t))/(2-lmb))

for(b in 1:length(shift)) {

mu = muy + shift[b]*(sigy/sqrt(n)) ; rlen = c()

for(i in 1:sim){

wt = ybar_Std = double(); cnt = 0; rl = 1

y = rnorm(n, mu, sigy); ybar_Std[1] = mean(y)

```

```

wt[1] = lmb * ybar_Std[1] + (1-lmb)*mu0

if(wt[1] > UCL[1] | wt[1] < LCL[1]) {cnt = 1}

while(cnt < 1){

  rl = rl + 1

  y = rnorm(n, mu, sigy);  ybar_Std[rl] = mean(y)

  wt[rl] = lmb * ybar_Std[rl] + (1-lmb)*wt[rl-1]

  if(wt[rl] > UCL[rl] | wt[rl] < LCL[rl]) {cnt = 1}

}

rlen[i] = rl

}

rlt[,b]=  rlen

arl[b] = mean( rlen)

sdt[b]=sd( rlen)

mdn[b]=median( rlen)

perc25[b]=quantile( rlen,0.25)

perc75[b]=quantile( rlen,0.75)

perc90[b]=quantile( rlen,0.90)

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

yn= rlt

list(xn,yn)

}

```

```
sim1=50000 ;muy1=0; sigy1=1 ;n1=1 ;lmb1=0.03; L1=2.483
```

```
w1_03=EWMA(muy=muy1,sigy=sigy1,lmb=lmb1,L=L1,sim=sim1, n=n1)[[1]]
```

Classical HWMA chart

```
HWMA=function(muy,sigy ,lmb,L,sim, n){

  shift= c(0, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.5, 0.75, 1, 1.5, 2)

  rlt=matrix(0,sim,length(shift));arl=c();sdt=c() ; mdn=c(); perc25=c(); perc75=c(); perc90=c()

  CL = UCL = LCL= double()

  ##### Regresion parameters

  mu0=muy; s0 = sqrt((sigy^2/n))

  LCL[1] = mu0 - L*s0*sqrt(lmb^2);

  UCL[1] = mu0 + L*s0*sqrt(lmb^2)

  t = 2:1e5

  LCL[t] = mu0 - L * s0*sqrt(lmb^2+(1-lmb)^2/((t-1)))

  UCL[t] = mu0 + L * s0*sqrt(lmb^2+(1-lmb)^2/((t-1)))

  for(b in 1:length(shift)) {

    mu = muy + shift[b]*(sigy/sqrt(n))

    rlen = c()

    for(i in 1:sim){

      wt = ybar_Std = double(); cnt = 0; rl = 1

      y = rnorm(n, mu, sigy); ybar_Std[1] = mean(y)

      wt[1] = lmb * ybar_Std[1] + (1-lmb)*mu0

      if(wt[1] > UCL[1] | wt[1] < LCL[1]) {cnt = 1}

      while(cnt < 1){
```



```

    rl = rl + 1

    y = rnorm(n, mu, sigy);    ybar_Std[rl] = mean(y)

    wt[rl] = lmb * ybar_Std[rl] + (1-lmb)*mean(ybar_Std[-rl])

    if(wt[rl] > UCL[rl] | wt[rl] < LCL[rl]) {cnt = 1}

}

rlen[i] = rl

}

rlt[,b]=  rlen

arl[b] = mean( rlen)

sdt[b]=sd( rlen)

mdn[b]=median( rlen)

perc25[b]=quantile( rlen,0.25)

perc75[b]=quantile( rlen,0.75)

perc90[b]=quantile( rlen,0.90)

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

yn= rlt

list(xn,yn)

}

sim1=50000 ;muy1=0; sigy1=1 ;n1=1 ;lmb1=0.03; L1=2.272

w1_03=HWMA(muy=muy1,sigy=sigy1,lmb=lmb1,L=L1,sim=sim1, n=n1)[[1]]

```

A_{ux} CUSUM₂ chart

```
library(MASS); library(spc)

MxCUSUM=function(K,H,sigy, muy, n, rho,sigx,mux,sim){

  shift= c(0, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.5, 0.75, 1, 1.5, 2)

  ybar= c(); covxy=rho*sigy*sigx; beta=covxy/sigx^2

  rlt=matrix(0,sim,length(shift));arl=c();sdt=c() ; mdn=c(); perc25=c(); perc75=c(); perc90=c()

  mup=muy; sdp = sqrt((sigy^2/n)*(1-rho^2))

  for(b in 1:length(shift)) {

    rlen=c()

    for(j in 1:sim)

    {

      m1= muy + shift[b]*(sigy/sqrt(n)) ; c1 = c2 =c()

      for(i in 1:100000)

      {

        z1 = rnorm(n); z2 = rnorm(n)

        x = sqrt(1-rho^2)*sigx*z1 + rho*sigx*z2 + mux; y = sigy*z2 + m1

        ybar = mean(y)+ beta*(mux-mean(x)) ; ybar= (ybar-mup)/sdp

        if(i==1)

        {

          c1[i] = max(0,(ybar-0) - K + 0)

          c2[i] = max(0,-(ybar-0) - K + 0)

        }

        else

        {
```

```

c1[i] = max(0,(ybar-0) - K + c1[i-1])

c2[i] = max(0,-(ybar-0) - K + c2[i-1])

}

if(c1[i] > H || c2[i]>H)

{

  rlen[j]=i

  #print(cbind(length(r1),r1[j],mean(r1)))

  break

}

else

{

  rlen[j]=100000

}

}

}

rlt[,b]= rlen

arl[b] = mean( rlen)

sdt[b]=sd( rlen)

mdn[b]=median( rlen)

perc25[b]=quantile( rlen,0.25)

perc75[b]=quantile( rlen,0.75)

perc90[b]=quantile( rlen,0.90)

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

```

```

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

yn= rlt

list(xn,yn)

}

sim1=50000 ; muy1=0; sigy1=1 ; mux1=0; sigx1=1 ;rho1=0.05 ;n1=10

K1=0.125; H1=xcusum.crit(k=K1, L0=500, mu0 = 0, hs = 0, sided = "two", r = 30)

```

M_X WWMA chart

```

M$_X$EWMA=function(muy,sigy,mux,sigx, rho, lmb,L,sim, n){

  shift= c(0, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.5, 0.75, 1, 1.5, 2)

  rlt=matrix(0,sim,length(shift));arl=c();sdt=c() ; mdn=c(); perc25=c(); perc75=c(); perc90=c()

  CL = UCL = LCL= double() ; mu0=muy; s0 = sqrt((sigy^2/n)*(1-rho^2))

  covxy=rho*sigy*sigx; beta=covxy/sigx^2

  t = 1:1e5

  LCL[t] = mu0 - L * s0*sqrt(lmb*(1-(1-lmb)^(2*t))/(2-lmb))

  UCL[t] = mu0 + L * s0*sqrt(lmb*(1-(1-lmb)^(2*t))/(2-lmb))

  for(b in 1:length(shift)) {

    mu = muy + shift[b]*(sigy/sqrt(n))

    rlen = c()

    for(i in 1:sim){

      wt = ybar_Std = double(); cnt = 0; rl = 1

      z1 = rnorm(n); z2 = rnorm(n)

      x = sqrt(1-rho^2)*sigx*z1 + rho*sigx*z2 + mux ; y = sigy*z2 + mu

```

```

ybar_Std[1] = mean(y)+ beta*(mux-mean(x)) ## regression equation

wt[1] = lmb * ybar_Std[1] + (1-lmb)*mu0

if(wt[1] > UCL[1] | wt[1] < LCL[1]) {cnt = 1}

while(cnt < 1){

  rl = rl + 1

  z1 = rnorm(n); z2 = rnorm(n)

  x = sqrt(1-rho^2)*sigx*z1 + rho*sigx*z2 + mux ; y = sigy*z2 + mu

  ybar_Std[rl] = mean(y)+ beta*(mux-mean(x)) ## regression equation

  wt[rl] = lmb * ybar_Std[rl] + (1-lmb)*wt[rl-1]

  if(wt[rl] > UCL[rl] | wt[rl] < LCL[rl]) {cnt = 1}

}

rlen[i] = rl

}

rlt[,b]= rlen

arl[b] = mean( rlen)

sdt[b]=sd( rlen)

mdn[b]=median( rlen)

perc25[b]=quantile( rlen,0.25)

perc75[b]=quantile( rlen,0.75)

perc90[b]=quantile( rlen,0.90)

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

```

```

yn= rlt

list(xn,yn)

}

sim1=50000 ;muy1=0; sigy1=1 ; mux1=0; sigx1=1 ;rho1=0.05;  n1=10

lmb1=0.03; L1=2.483

w1_03=M$_X$EWMA(muy=muy1,sigy=sigy1,mux=mux1,sigx=sigx1, rho=rho1,lmb=lmb1,L=L1,sim=sim1, n=n1)[[

```

HWMA_Y chart

```

HWMA_Y=function(muy,sigy,mux,sigx, rho, lmb,L,sim, n){

shift=  c(0, 0.05, 0.075, 0.1, 0.125, 0.15, 0.175, 0.2, 0.25, 0.5, 0.75, 1, 1.5, 2)

rlt=matrix(0,sim,length(shift));arl=c();sdt=c() ;  mdn=c(); perc25=c(); perc75=c(); perc90=c()

CL = UCL = LCL= double()

##### Regresion parameters

mu0=muy; s0 = sqrt((sigy^2/n)*(1-rho^2))

covxy=rho*sigy*sigx;  beta=covxy/sigx^2

##### UCLlimit

LCL[1] = mu0 - L*s0*sqrt(lmb^2);

UCL[1] = mu0 + L*s0*sqrt(lmb^2)

t = 2:1e5

LCL[t] = mu0 - L * s0*sqrt(lmb^2+(1-lmb)^2/((t-1)))

UCL[t] = mu0 + L * s0*sqrt(lmb^2+(1-lmb)^2/((t-1)))

for(b in 1:length(shift)) {

mu =  muy + shift[b]*(sigy/sqrt(n))

rlen = c()

```

```

for(i in 1:sim){

  wt = ybar_Std = double(); cnt = 0; rl = 1

  z1 = rnorm(n); z2 = rnorm(n)

  x = sqrt(1-rho^2)*sigx*z1 + rho*sigx*z2 + mux ; y = sigy*z2 + mu

  ybar_Std[1] = mean(y)+ beta*(mux-mean(x)) ## regression equation

  wt[1] = lmb * ybar_Std[1] + (1-lmb)*mu0

  if(wt[1] > UCL[1] | wt[1] < LCL[1]) {cnt = 1}

  while(cnt < 1){

    rl = rl + 1

    z1 = rnorm(n); z2 = rnorm(n)

    x = sqrt(1-rho^2)*sigx*z1 + rho*sigx*z2 + mux ; y = sigy*z2 + mu

    ybar_Std[rl] = mean(y)+ beta*(mux-mean(x)) ## regression equation

    wt[rl] = lmb * ybar_Std[rl] + (1-lmb)*mean(ybar_Std[-rl])

    if(wt[rl] > UCL[rl] | wt[rl] < LCL[rl]) {cnt = 1}

  }

  rlen[i] = rl

}

rlt[,b]= rlen

arl[b] = mean( rlen)

sdt[b]=sd( rlen)

mdn[b]=median( rlen)

perc25[b]=quantile( rlen,0.25)

perc75[b]=quantile( rlen,0.75)

perc90[b]=quantile( rlen,0.90)

```

```

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

yn= rlt

list(xn,yn)

}

sim1=50000;muy1=0; sigy1=1 ; mux1=0; sigx1=1 ;rho1=0.05;n1=1

lmb1=0.03; L1=2.272

w1_03=classical_HWMA(muy=muy1,sigy=sigy1,mux=mux1,sigx=sigx1, rho=rho1,

lmb=lmb1,L=L1,sim=sim1, n=n1)[[1]]

```

Corrected limit of the AHWMA chart

```

library(spc)

AHWMA=function(muz,sigz,muy,sigy, rho, w,L,sim, n){

  shift= c(0)

  rlt=matrix(0,sim,length(shift));arl=c();sdt=c() ; mdn=c(); perc25=c(); perc75=c(); perc90=c()

  CL = UCL = LCL= double()

  ##### Regresion parameters

  mu0=muz; s0 = sqrt((sigz^2/n)*(1-rho^2))

  covxy=rho*sigz*sigy; beta=covxy/sigy^2

  #####

  LCL[1] = mu0 - L*s0*sqrt(w^2);

  UCL[1] = mu0 + L*s0*sqrt(w^2)

```



```

t = 2:1e5

LCL[t] = mu0 - L * s0*sqrt(w^2+(1-w)^2/((t-1)))

UCL[t] = mu0 + L * s0*sqrt(w^2+(1-w)^2/((t-1)))

for(b in 1:length(shift)) {

  mu = muz + shift[b]*(sigz/sqrt(n))

  rlen = c()

  for(i in 1:sim){

    wt = ybar_Std = double(); cnt = 0; rl = 1

    z1 = rnorm(n); z2 = rnorm(n)

    x = sqrt(1-rho^2)*sigy*z1 + rho*sigy*z2 + muy ;

    y = sigz*z2 + mu

    ybar_Std[1] = mean(y)+ beta*(muy-mean(x)) ## regression equation

    wt[1] = w * ybar_Std[1] + (1-w)*mu0

    if(wt[1] > UCL[1] | wt[1] < LCL[1]) {cnt = 1}

    while(cnt < 1){

      rl = rl + 1

      z1 = rnorm(n); z2 = rnorm(n)

      x = sqrt(1-rho^2)*sigy*z1 + rho*sigy*z2 + muy ; y = sigz*z2 + mu

      ybar_Std[rl] = mean(y)+ beta*(muy-mean(x)) ## regression equation

      wt[rl] = w * ybar_Std[rl] + (1-w)*mean(ybar_Std[-rl])

      if(wt[rl] > UCL[rl] | wt[rl] < LCL[rl]) {cnt = 1}

    }

    rlen[i] = rl

  }

}

```

```

rlt[,b]= rlen

arl[b] = mean( rlen)

sdt[b]=sd( rlen)

mdn[b]=median( rlen)

perc25[b]=quantile( rlen,0.25)

perc75[b]=quantile( rlen,0.75)

perc90[b]=quantile( rlen,0.90)

}

xn=matrix(round(cbind(arl,sdt, mdn,perc25, perc75, perc90),2),length(shift),6)

colnames(xn)=c("ARL", "SDARL", "MDRL", "Per25", "Per75", "Per90")

rownames(xn)=shift

yn= rlt

list(xn,yn)

}

```

```

BiSearch <- function(ARL, muz,sigz,muy,sigy, rho, w, sim, n) {

h2 = sapply(c(ARL), l=w, sided="two",xewma.crit) - 0.5

ARL2 = AHWMA(muz,sigz,muy,sigy, rho, w, h2, sim, n)[[1]][1]

while(ARL2 < ARL){

h1 = h2;      ARL1 = ARL2;      h2 =h2+ 1.

ARL2 = AHWMA(muz,sigz,muy,sigy, rho, w,h2,sim, n)[[1]][1]

}

h3 = h1 + (ARL-ARL1)/(ARL2-ARL1) * (h2-h1)

```

```

ARL3 = AHWMA(muz,sigz,muy,sigy, rho, w,h3,sim, n)[[1]][1]

dc = h3 - h2;h1 = h2; ARL1 = ARL2; h2 = h3; ARL2 = ARL3;

while((abs(ARL-ARL3)> 0.5)){

  h3 = h1 + (ARL-ARL1)/(ARL2-ARL1) * (h2-h1);

  ARL3 = AHWMA(muz,sigz,muy,sigy, rho, w,h3,sim, n)[[1]][1];

  h1 = h2; ARL1 = ARL2; h2 = h3; ARL2 = ARL3

}

xn=matrix(round(cbind(w, h3, ARL3),4),1,3)

colnames(xn)=c("smoothing parameter", " C", "ARL")

return(xn)

}

ARL=500 ## the in-control Average run length (ARL).

muz=0; sigz=1 # the mean and sigma parameters of the process variable of interest

muy=0; sigy=1 # the mean and sigma parameters of the auxiliary variable

rho=0.05 # the correlation size

w=0.03 # the smoothing parameter-- this need to be changed to the required value

sim=500 #the simulation size use to obtain the value of the ARL.

#We recommend that the "sim" should be large for

#correctness (sim size should not be less than 50,000).

C=BiSearch(ARL, muz,sigz,muy,sigy, rho, w, sim, n=1) # this gives the width of the

#proposed chart when the smoothing parameter is 0.03

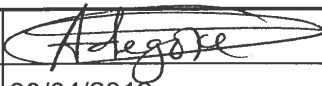
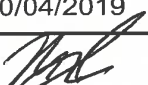
```




MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
Shrinkage estimates of covariance matrices to improve the performance of multivariate cumulative sum control charts. Computers & Industrial Engineering, 117, 207-216.	
In which Chapter is the Manuscript /Published work:	Chapter 2
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	50%
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	
Participated in the formulation and studying of the proposed solution. Conducted literature search, and wrote the first draft of the manuscript.	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	1/5/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
Multivariate exponentially weighted moving average (MEWMA) control charts for individual-observations monitoring when parameters are estimated	
In which Chapter is the Manuscript /Published work:	Chapter 3
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	70%
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	
<p>participated in the formulation and studying of the proposed solution. Conducted literature search and wrote the first draft of the manuscript</p>	
For manuscripts intended for publication please indicate target journal:	
Submitted to the Journal of Applied Statistics	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	1/5/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
Multivariate coefficient of variation control charts in phase I of SPC. The International Journal of Advanced Manufacturing Technology, 99(5-8), 1903-1916.	
In which Chapter is the Manuscript /Published work:	Chapter 4
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	25%
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	
Participated in the formulation and studying of the proposed solution. Literature searched and wrote the first draft of the manuscript.	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	1/5/2019



(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
A Multivariate Homogeneously Weighted Moving Average Control Chart. IEEE Access, 7, 9586-9597.	
In which Chapter is the Manuscript /Published work:	Chapter 5
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	85%
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	
<p>Participated in the formulation and studying of the proposed solution. Conducted literature search and wrote the first draft of the manuscript.</p>	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	1/5/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
A new distribution-free multivariate control chart for ecological applications	
In which Chapter is the Manuscript /Published work:	Chapter 6
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	65 %
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	
<p style="font-family: cursive;">Participated in the formulation and studying of the proposed solution. conducted literature search and wrote the first draft of the manuscript</p>	
For manuscripts intended for publication please indicate target journal:	
Ecology	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	17/5/2019

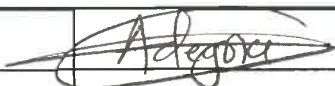
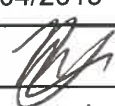
(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
EWMA-type scheme for monitoring location parameter using auxiliary information. Computers & Industrial Engineering, 114, 114-129.	
In which Chapter is the Manuscript /Published work:	Chapter 7
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	85%
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	
<p>participated in the formulation and studying of the proposed solution. Conducted literature search and wrote the first draft of the manuscript.</p>	
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	1/5/2019

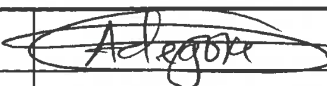
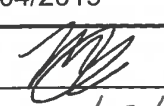
(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
Enhancing the performance of the EWMA control chart for monitoring the process mean, using auxiliary information. Quality and Reliability Engineering International.	
In which Chapter is the Manuscript /Published work:	Chapter 8
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	80%
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	<p>participated in the formulation and studying of the proposed method. conducted literature search and wrote the first draft of the manuscript.</p>
For manuscripts intended for publication please indicate target journal:	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	1/5/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)



MASSEY UNIVERSITY
GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Nurudeen Adedayo Adegoke
Name/title of Primary Supervisor:	Dr. Matthew D.M. Pawley
Name of Research Output and full reference:	
Efficient monitoring of a process mean using an auxiliary variable under simple random sampling	
In which Chapter is the Manuscript /Published work:	Chapter 9
Please indicate:	
<ul style="list-style-type: none"> The percentage of the manuscript/Published Work that was contributed by the candidate: 	80%
and	
<ul style="list-style-type: none"> Describe the contribution that the candidate has made to the Manuscript/Published Work: 	<p>Participated in the formulation and studying of the proposed solution. Conducted literature search and wrote the first draft of the manuscript.</p>
For manuscripts intended for publication please indicate target journal:	
Submitted to IEEE Access	
Candidate's Signature:	
Date:	30/04/2019
Primary Supervisor's Signature:	
Date:	1/5/2019

(This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis)